

Handle with care

Scientists are rushing to defend a colleague charged with mishandling samples of the plague bacterium. But they must be careful not to send the message that microbiologists are blasé about the need to protect public health.

When Thomas Butler stepped off a plane in April 2002 on his return to the United States from a trip to Tanzania, he set in motion a chain of events that now threatens to destroy his life. A microbiologist at Texas Tech University in Lubbock, Butler was bringing back samples of the plague bacterium *Yersinia pestis* for his research. Yet on re-entering the country, he is alleged to have passed right by US customs inspectors without notifying them that he was carrying this potentially deadly cargo.

That move and its repercussions have led the federal government to prosecute Butler for a range of offences. He was arrested in January after declaring to federal officials that plague samples had gone missing. Butler then admitted — under duress, he now claims — that he had destroyed the samples without following federal guidelines for handling such dangerous materials. His trial began in Lubbock on 3 November. If convicted on all counts, he could be fined millions of dollars and spend the rest of his life in jail.

The US scientific community has leapt to Butler's defence, arguing that his prosecution is overzealous, alarming and unnecessary. The presidents of the National Academy of Sciences and the Institute of Medicine have written to Attorney General John Ashcroft, claiming that the case could endanger research into countering the threat of bioterrorism. And the academy's human-rights committee has asked its members to write letters on Butler's behalf and to donate funds for his defence (see *Nature* **425**, 5; 2003).

Those who defend Butler argue that the rules governing the import of pathogens are so restrictive that bending them is the only option for researchers who are working to provide protection from deadly diseases that afflict the developing world (see *Nature* **423**, 669; 2003). Why, they ask, prosecute Butler for infractions of rules that made his work more difficult without serving a useful purpose?

The charges laid against Butler run further than his alleged flouting of import controls, however. He is accused of having travelled across the United States with plague bacteria at least twice last year, and of shipping plague samples back to Tanzania with a label identifying the contents merely as 'lab materials'. He is also charged with inflating research expenses to avoid \$40,000 in taxes.

His supporters counter that this damning charge sheet merely reflects the determination of federal prosecutors to throw the book at Butler to make an example of him to others. Many researchers now fear falling victim to an overzealous prosecution if they fail to dot all the 'i's and cross all the 't's on their paperwork. Some US microbiologists are so frightened of being hauled off in handcuffs for a minor administrative oversight that they have decided to avoid biodefence research entirely — despite the current funding bonanza in the field.

Whether Butler is a villain or a scapegoat is now for a jury to decide. But whatever verdict is eventually reached, scientists who are lobbying on Butler's behalf would do well to consider public perceptions. If the rules governing the import of pathogenic bacteria make no sense, then microbiologists must make that case clearly, and lobby for the regulations to be changed. Researchers are also justified in making representations to help ensure that any punishment that Butler might receive is proportionate.

But researchers risk a damaging public backlash if the main message that emerges is that his peers think he was justified in carrying samples of the plague bacterium onto a commercial flight. Appearing to deny the importance of rules designed to protect the public from deadly pathogens — however unwieldy those rules may be in practice — will not engender trust. It will not foster a culture of responsibility. And it would show disregard for the public's faith that scientific research will be conducted as safely and as competently as possible. ■

Defamation, online

Operators of preprint archives and other scientific websites would be well advised to get up to speed on media law.

Regular users of the ArXiv physics preprint server know that its contents occasionally depart from the dry format of the standard scientific manuscript — recent examples include an essay that draws parallels between astrophysics and prostitution (M. López-Corredoira, preprint at <http://arxiv.org/astro-ph/0310368>; 2003). For the most part, such contributions serve as mild irritations or amusing distractions, depending on your point of view.

But a posting made on 27 October (A. De Rújula, preprint at <http://arxiv.org/physics/0310134>; 2003) raises more serious concerns. The preprint accuses Britain's astronomer royal of claiming credit for other researchers' ideas. And legal experts contacted by *Nature* argue that the language used could, under some jurisdictions, be considered libellous (see page 7).

Magazines, newspapers and some scientific journals consult specialists in media law before publishing articles that are potentially defamatory. They also take out insurance to cover the eventuality of losing a libel suit. But the operators of preprint archives have not,

until now, seen the need to take such precautions. Paul Ginsparg of Cornell University, who runs ArXiv, argues that a preprint server cannot be judged by the standards of a newspaper.

In some respects, he is correct — under many jurisdictions, the operators of a website won't necessarily be assumed to have read and approved every posting made. But in the event of a complaint being made, things get complicated. The Internet's international reach means that a website's operators can, in theory, be held liable under the law of any country in which the material is accessed. In the case of a potentially libellous posting, that could mean weighing the probable liability under England's tough defamation laws, if the posting were not removed, with a US author's right to free speech, enshrined under the First Amendment to the constitution.

Thankfully, most scientists are not especially litigious. But losing just one legal action can cost an awful lot of money. If De Rújula's posting marks the start of a trend, every operator of a preprint server may need to find a good lawyer. ■





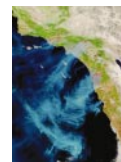
Sunshine deal
Critics question
Scripps Florida
agreement
p4



Bearing the brunt
Long-term pollutants
cause problems for
polar bears
p5



Deep thoughts
Call for US action
to explore new
ocean regions
p6



Fire breaks
San Diego blaze
forces labs to delay
grant applications
p8

Satellite loss throws Japan's space programme into disarray

David Cyranoski, Tokyo

Japan's space agency is facing tough questions this week after the unexplained loss of Midori-II, one of the nation's most ambitious scientific satellites to date.

Shuichiro Yamanouchi, president of the newly revamped Japan Aerospace Exploration Agency (JAXA), declared the Earth observation probe lost for good on 31 October, a week after it stopped sending signals home. "It's regrettable," Yamanouchi said of the loss of the ten-month-old mission. "At this point there's little more to say."

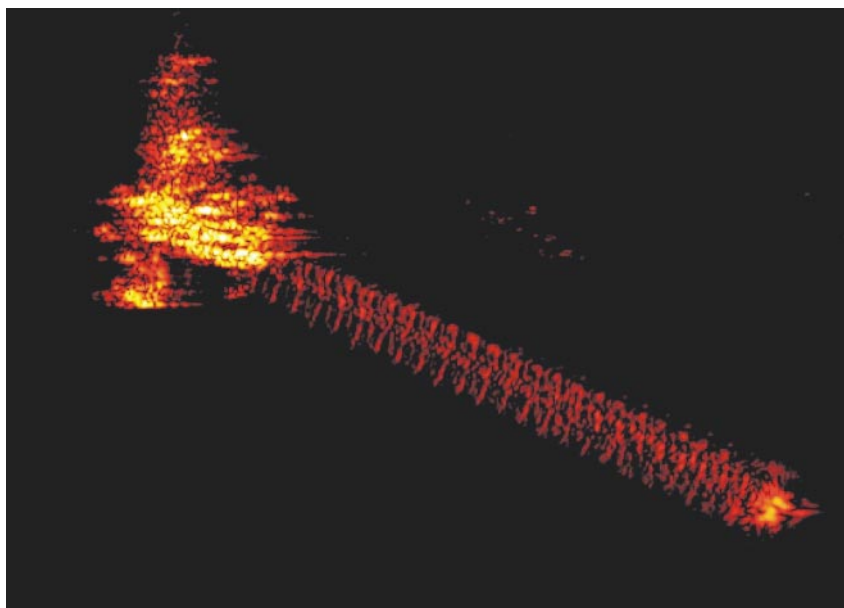
But the failure of the ¥70-billion (US\$630-million) project, which was carrying instruments for Japan's space partners Europe and the United States, is a heavy blow for the space agency, which was reconstituted only last month in a bid to boost the national space programme.

This is the second such loss in six years. In June 1997, Midori, a Japanese Earth observation satellite carrying French and US instruments, was lost after ten months of operation. And following a costly, decade-long struggle to establish a reliable rocket-launching capacity, JAXA can scarcely afford this latest setback.

Midori-II was carrying five instruments for studying various oceanic, terrestrial and atmospheric features. They included an upgrade of a NASA instrument for measuring the oceans' interaction with climate that was lost with Midori. And the CNES, France's national space agency, had sent an identical version of its POLDER sensor to that on the previous mission. The sensor measures solar radiation from aerosols, clouds, the oceans and land surfaces.

Midori-II had already sent back important data for climate studies, its backers say. These included data on soil moisture and sea surface temperature obtained by JAXA's Advanced Microwave Scanning Radiometer from measurements of microwave emissions from Earth's surface and atmosphere.

But on 25 October, Midori-II began to miss data-transmission appointments. The satellite has not responded to signals since. It was later discovered that power generation from its solar battery paddle had fallen sharply



Radar images of troubled satellite Midori-II show that the solar panels powering the craft remain intact.

six hours earlier, from 6 kW to 1 kW. JAXA is looking into the cause of the power loss.

The satellite was last observed in radar images taken by the Research Institute for High Frequency Physics and Radar Techniques in Wachtberg, Germany. They showed that the solar panels that powered the craft were still intact. Its predecessor's loss was directly attributed to damage to the solar panels.

JAXA is now under pressure to explain Midori-II's power problems. Initial reports blamed a solar flare — a sudden burst of energy from the Sun's surface — that sent a stream of charged particles into the atmosphere over Peru at the time Midori-II was passing. The timing coincided with the abrupt drop in power. But JAXA officials say they are not definitely blaming the flare.

"The solar flare is one consideration, but at this point we cannot say whether or not it had any relation," JAXA's associate executive director Tsuguhiko Katagi said at the 31 October press conference. JAXA has 100 researchers looking into the power failure. The investigation, led by Yamanouchi, will

report to the education ministry's Space Activities Commission. But JAXA officials say that the inquiry could, as in the case of Midori, take up to six months.

Yamanouchi deflected questions about how JAXA might salvage its reputation, or how the gap left in Japanese Earth and climate studies would now be filled. "This research is a matter of international collaboration and will need to be discussed," he said.

"It was a real shocker for me," says Timothy Liu, the SeaWinds project director at NASA's Jet Propulsion Laboratory in Pasadena, California. The SeaWinds scatterometer for measuring ocean wind speeds was also lost on Midori-II. After the loss of its first device in 1997, Liu's lab used spare parts for fast assembly of a replacement. Liu says it has been "very disappointing to lose both satellites" and adds that they have no money for a replacement. He does not rule out any future collaboration with Japan's space programme. "But we hope that by that time Japan has improved its assessment and response measures," he adds. ■

► <http://sharaku.eorc.jaxa.jp/ADEOS2/sensor/sensor.html>

Disease spread leads Britain to pull plug on trial badger cull

Jim Giles, London

A strategy to control tuberculosis (TB) in cattle by killing thousands of badgers a year seems to do more harm than good, according to the initial results of a nationwide study in Britain.

Badgers carry the bacterium that causes TB in cows, and are thought to be an important reservoir for the disease. In an attempt to control bovine TB, the British government has overseen the culling of more than 20,000 badgers since 1975 in areas where herds are found to be infected — a strategy known as 'reactive culling'. This policy was put on hold in 1998 so that a study could assess the effectiveness of the method.

That study, led by John Bourne, a retired veterinarian formerly at the University of Reading, compared reactive culling with two other methods: no culling and 'proactive culling', in which badger numbers are reduced across entire trial areas regardless of the incidence of TB. The study involved 30 regions of England, each around 100 km² in size, and was planned to run until 2005.

But Bourne's team decided last month to notify the government of the trial's initial results, as the number of infected herds in reactive-culling areas is now 27% higher than in regions without culls. The study does not reveal the cause of the increase, but researchers think that the culling may have prompted badgers to move between territories, increasing the spread of the disease. More results are required before they can judge whether the proactive strategy is more, or less, effective.

Ben Bradshaw, Britain's minister for nature conservation and fisheries, announced on 4 November that reactive culling will no longer be used in the trial. The results are disappointing for some, says Bradshaw, as reactive culling resulted in the death of fewer badgers than the proactive culling used in the trial. If effective, it could have been a compromise between farmers, who want to protect their cattle, and animal-welfare groups who oppose culling. "The reactive option would have been one of the most attractive policies for many people," he says.

Stephen Harris, an expert on TB and badgers at the University of Bristol, notes that proactive culling can be altered to have a lesser effect on badger populations and the environment. But in the end, he notes, "the only long term solution to the problem is to develop a cattle vaccine". ■

Florida welcomes Scripps as critics predict big-money flop

Rex Dalton, San Diego

Florida governor Jeb Bush this week signed legislation that will help to pump \$570 million into bringing the Scripps Research Institute to the 'sunshine state'. But critics are already asking if the deal will attract the research jobs and biotechnology companies that its proponents claim.

The initiative to create Scripps Florida at Palm Beach has been spearheaded by Bush — brother of President George Bush. Last month, he called a special session of the state legislature to put \$310 million into the deal, to pay for operations as the new institute builds up to house a projected 545 researchers after seven years.

On 23 October, the legislature approved the request, which includes another \$60 million in projected investment income from money being put aside now to fund operations. Palm Beach County will also pay \$200 million to build a huge lab complex for the institute. Bush has said the deal could bring up to 500 new firms and 44,000 jobs to Florida.

But sceptical Florida law-makers and economists say that this greatly overstates the benefits of the deal with Scripps, a non-profit organization whose only facility is in La Jolla, California. They also point to unanswered questions about the possible role of the Swiss drug firm Novartis in Scripps Florida's future.

Novartis and Scripps have a contract that says the pharmaceutical firm has first rights to select for product development up to 47% of ideas from the research institute at La Jolla. Novartis is paying Scripps \$20 million a year under this pact, which created considerable controversy when it was signed in 1994 by Scripps and the Swiss-based drug company Sandoz — since incorporated into Novartis. Critics of the deal said that Scripps,



Jeb Bush signs the deal he hopes will reap biotech riches.

which gets 85% of its income from the National Institutes of Health, should not have allocated such broad product-development rights to one company.

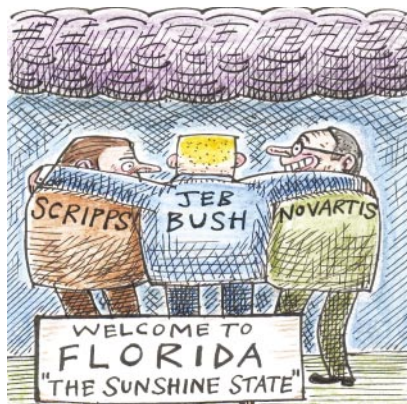
Neither Novartis nor Scripps will say whether the Florida facility will be covered by the contractual agreement, which expires in December 2006. A Scripps spokesman says that the concept of Novartis selecting technology developed at Scripps Florida "is under discussion". Novartis officials refused repeated requests for comment.

In Florida, Jill Bratina, a spokeswoman for Bush, says that the existing Novartis-Scripps contract won't cover Scripps Florida. But she also adds that the governor's office is in discussion with Novartis and Novartis-funded researchers in La Jolla about possible future involvement in Scripps Florida.

If a renewed Novartis contract with Scripps did cover the Florida facility, it could have a major impact on the number of companies and biotechnology jobs that would spin off from it, critics say. Novartis would get automatic access to the most promising ideas for commercial development, they note.

Dan Gelber, a Democrat state representative for Miami Beach, who voted against the Scripps package in the legislature, says that most of his colleagues were unaware of such criticisms. "There was no deep thought put into this," he says. "The legislature didn't do the due diligence the taxpayers would hope for; it was the equivalent of buying a used car."

Gelber says independent experts estimate that the deal will create only 16,000 jobs over 15 years. And Joseph Cortright, an economist in Portland, Oregon, who specializes in the impact of biotechnology on regional development, calls the estimate of 500 new firms "ludicrous". Bratina counters that the governor's economic forecasts are solidly based on "conservative" estimates. ■



US loses allure in foreign students' eyes

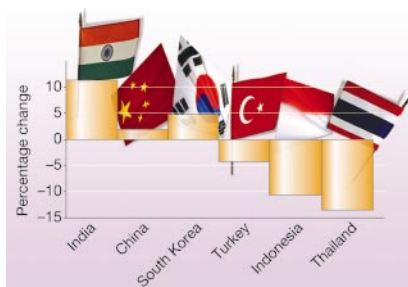
Geoff Brumfiel, Washington

The number of foreign students enrolling in US universities stagnated last year and may now be declining for the first time in decades, according to a series of studies released this week.

The annual report of the Institute of International Education (IIE), a New York-based organization that monitors student exchange, shows that the growth in total foreign-student enrolment at US universities grew by just 0.6% in the 2002–03 academic year — the lowest level of growth since 1995. And in a follow-on survey of international student offices in 276 universities, most reported some decline in international enrolment this autumn, suggesting that the total number of students may actually fall in the 2003–04 academic year for the first time since 1971.

Such a drop is likely to hit US universities where it hurts: international students spend some US\$12 billion annually on tuition and living expenses, according to the IIE, and foreign graduate students fill many research and teaching jobs on campuses, particularly in science and engineering.

The 2002–03 academic year saw sharp



Learning curve: the changing numbers of foreign students coming to the US between 2002 and 2003.

increases in the number of students coming from India and South Korea, the IIE survey found, but the number arriving from Europe, Southeast Asia and the Middle East declined.

University administrators and academic groups attribute the slowing enrolment rates largely to the strict visa-issuance policies enacted in the wake of the terrorist attacks of 11 September 2001 (see *Nature* **422**, 457–258; 2003). “The US government is doing more than anyone to deter foreign students,” says Gail Szenes, director of the Office for International Students and Scholars at

New York University, whose campus lies just blocks from the World Trade Center, destroyed in the 2001 terror attacks. Consular interviews for students whose subject areas appear on a state department watch list are especially problematic for those studying in the sciences, Szenes says.

Visa woes are also afflicting scholars coming to teach and to do research in the United States, according to preliminary results of a separate survey by three major university associations. About three-quarters of more than 200 US universities surveyed by the associations said that they had problems getting scholars into the country, according to Victor Johnson, public-policy director at the Washington DC-based Association of International Educators, one of the groups that conducted the survey.

New visa guidelines have created a perception that it is more difficult for foreign students and scholars to come to the United States, but until now, the evidence for this was largely anecdotal, according to Johnson. “These are the first data that confirm our fears,” he says. ■

► <http://opendoors.iienetwork.org>

► www.nafsa.org

Roaming polar bears reveal Arctic role of pollutants

Marika Willerroider, Munich

Female polar bears have provided researchers with an insight into the long-term health impact of pollutants.

In a study published on 1 November, a Norwegian team reveals that bears that roam long distances in their search for food tend to accumulate relatively high levels of industrial pollutants such as polychlorinated biphenyls (PCBs) in their bodies. This could be linked to the exceptional levels of hermaphroditism seen in some bear populations, although this connection has yet to be confirmed.

PCBs, which were widely used in electrical equipment until they were banned in most countries on health grounds, are extremely resistant to biodegradation. They accumulate in the food chain, even in remote locations, and then accrue at the chain's end — in polar bears, for example, which cannot easily metabolize or excrete the compounds.

The researchers tagged 54 female bears from the Svalbard archipelago and the nearby Barents Sea in the far north of Europe with satellite transmitters, and monitored their movements for several years. They found that some bears in off-



At the end of a food chain, polar bears accumulate toxic chemicals.

coast habitats roam huge territories each year — up to 270,000 square kilometres — and accumulate significantly higher levels of PCBs in their fat, blood and milk than bears in smaller coastal or near-coastal habitats (G. H. Olsen *et al. Environ. Sci. Technol.* **37**, 4919–4924; 2003).

“Long-distance migration costs immense energy,” says Øystein Wiig, a mammalogist at the University of Oslo's Zoological Museum and a co-author on the study. “The bears need to consume much more prey, and thus build up more PCBs.”

Particularly high pollutant concentrations were found in bears migrating to the Kara

Sea, north of Russia. This indicates that there is massive pollution of Siberian rivers from contaminated industrial areas upstream, says Wiig.

PCBs are part of a group of pollutants known as endocrine disruptors, which may interfere with hormone function in humans and animals at relatively low doses. Earlier research suggests that the PCBs could be the cause of sexual anomalies found in polar bears — hermaphroditism has been seen in about 2% of the roughly 5,000-strong bear population in

Svalbard (Ø. Wiig *et al. J. Wildlife Dis.* **34**, 792–796; 1998).

But the link is tentative. “I’d rather not jump to conclusions. Hermaphrodites have also been found among brown bears that are free of PCBs,” says Aaron Fisk, a toxicologist at the University of Georgia in Athens, who investigates pollution in arctic wildlife.

The release of PCBs into the arctic environment is slowly decreasing, says Samantha Smith, who directs the arctic programme of the WWF, an environmental group. But she warns that a second wave of pollutants, now stored in arctic ice reservoirs, may be released if glaciers melt. ■

Museum breaks mould in attempts to lure reluctant visitors

Jim Giles, London

Museum curators are fond of saying that people visit their exhibits only three times: as children, with their own children, and again with their grandchildren. Now Britain's Science Museum is confronting this sorry situation head-on, opening a £10-million (US\$17-million) centre aimed directly at adults.

The Dana Centre in South Kensington, London, will open its doors on 19 November and is targeted squarely at visitors aged 18–45. Events will range from evening discussions on AIDS in ethnic minority communities, to film premières, as well as giving visitors the chance to watch a live broadcast of a heart operation and quiz the surgeons involved.

The centre could also feature risqué happenings more associated with other adults-only venues. Dana Centre staff point out that the museum's stores contain Victorian sex toys, and the centre's child-free environment might make it possible to run events themed on such objects.

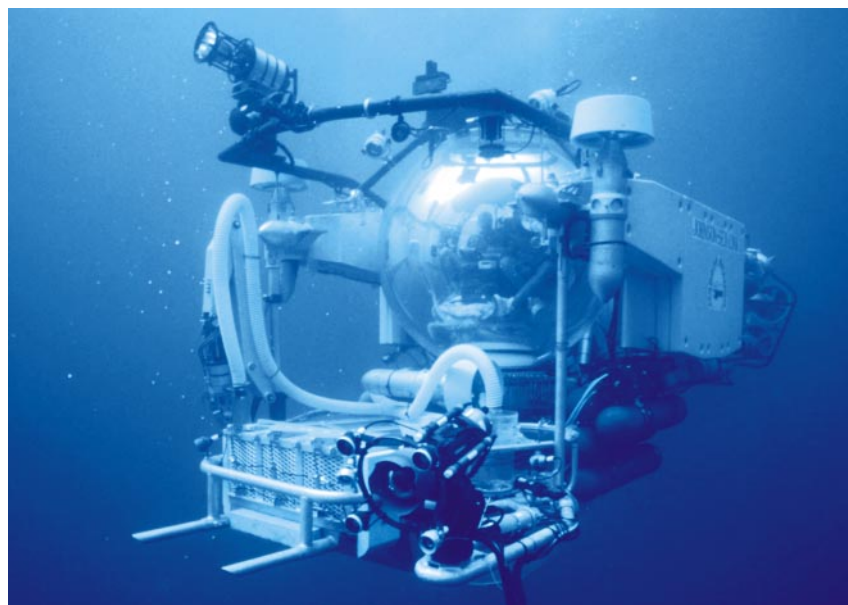
"It they pull this off it will be incredible," says Steve Cross, science-development manager at the Life Science Centre in Newcastle upon Tyne, UK. Adults aged 18–45 tend to have established interests, he points out, and it is tough to make them take a second look at science. "Everyone in the science-communication world is scared of that age range," says Cross. "If they succeed they will have rewritten the rules."

The centre may also struggle to achieve its aim of attracting visitors in the evening, sceptics say. South Kensington is certainly better known for its museums and embassies than for its nightlife.

But the centre's staff counter that such problems can be overcome by hosting carefully designed events. This February, for example, the museum successfully targeted the Afro-Caribbean community with a lecture timed to coincide with *Motherland*, a BBC television documentary in which DNA analysis was used to help British Afro-Caribbeans trace their ancestry back to specific parts of Africa.

As well as hosting science events based on stand-up comedy and contemporary dance, the centre hopes to build on the success of the Café Scientifique idea of holding events in bars and cafés at which visiting researchers discuss scientific topics. Such events already take place across Europe and the United States. ■

♦ www.danacentre.org.uk



Hidden depths: research missions are ignoring previously unexplored areas of the oceans.

National Academy calls for sea change in ocean efforts

Betsy Mason, Washington

The United States isn't doing enough to explore the oceans, and needs to develop an inter-agency approach to such exploration, says a committee of the National Academy of Sciences.

An academy panel says that a better understanding of the oceans requires a major, coordinated, international effort to investigate unexplored regions. And the best way to get such an effort under way is for the United States to take the lead, in the hope that other countries will follow suit, the panel concludes in a report released on 4 November.

Since 2000, the US government has sponsored a small ocean-exploration programme at the National Oceanic and Atmospheric Administration (NOAA), but limited funding has kept most projects from venturing beyond North American coastal waters.

Outside NOAA, most ocean research has focused on specific objectives that often involve further study of already known areas. Little money or effort has gone into pure exploration, and most discoveries occur serendipitously, says the panel.

"We need to encourage and nurture that type of serendipity," says biological oceanographer and panel member Shirley Pomponi of Harbor Branch Oceanographic Institution in Fort Pierce, Florida.

The only way to achieve this is to make funds available specifically for exploring new areas such as the Southern and Arctic oceans, says committee chair John Orcutt, a marine geophysicist at Scripps Institution of

Oceanography in La Jolla, California. The panel estimates that a high-quality programme, including a fleet of submersibles and research vessels, would need around \$270 million to set up and \$110 million in annual operating costs. A more modest programme that could still achieve many of the exploration goals would cost \$70 million per year. NOAA gets just \$14 million a year for its existing programme.

The panel is critical of NOAA's ability to handle a larger programme, however, and says that an inter-agency approach would be a better solution.

But James Baker, president of the Academy of Natural Sciences in Philadelphia and former NOAA head under the Clinton administration, thinks that the programme would do better inside NOAA or another existing agency. "Big initiatives always do best if they are single-agency efforts," says Baker. "We need to pick one and make it work."

Congressman Jim Greenwood (Republican, Pennsylvania), a member of the House Oceans Caucus who was instrumental in asking for the report, strongly endorsed its findings. "By not exploring the rest of the oceans, we are missing the opportunity to discover all kinds of pharmaceuticals and natural resources," he says.

Greenwood says he is optimistic that the Bush administration will look seriously at supporting ocean research, to boost its environmental reputation. "A robust and romantic approach to exploring the oceans would be good policy and good politics," he says. ■

Preprint server seeks way to halt plagiarists

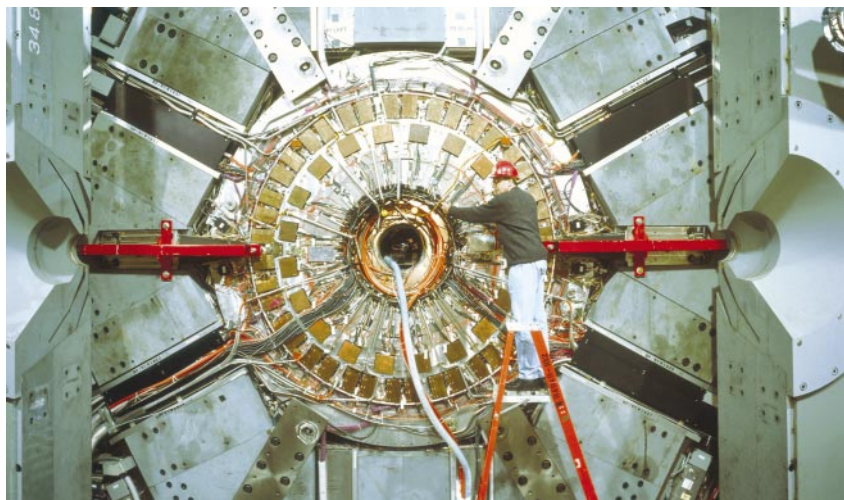
Jim Giles, London

An unusual case of plagiarism has struck ArXiv, the popular physics preprint server at Cornell University in Ithaca, New York, resulting in the withdrawal of 22 papers.

In the light of the incident, Cornell physicist Paul Ginsparg, who developed the archive, says that he now intends to investigate measures to prevent such misdemeanours from recurring. In the meantime, Ginsparg is also facing the challenge of potentially libellous material being posted on the server (see below).

The plagiarism case traces its origins to June 2002, when Yasushi Watanabe, a high-energy physicist at the Tokyo Institute of Technology, was contacted by Ramy Naboulsi, who said he was a mathematical physicist. Naboulsi asked for Watanabe's help in obtaining a research position in Japan. Impressed by Naboulsi's work, Watanabe agreed to upload some of his papers to ArXiv, which Naboulsi was unable to do himself as he had no academic affiliation. "I was so amazed at his productivity I began to think he was a genius," Watanabe later wrote in an e-mail to the archive.

By April 2003, Naboulsi had 22 papers on ArXiv, but some of server's users noticed that one of his papers copied parts of the *BaBar Physics Book*, an online summary of meetings about a high-energy physics experiment at the Stanford Linear Accelerator Center in California. When six more of the papers were shown to be similar to the BaBar book, Watanabe asked for all 22 preprints to be withdrawn. ArXiv labelled the papers as



Plundered: experimental results from the BaBar particle detector have been copied by an author.

such but left them on the site, in accordance with its policy.

When contacted by *Nature*, Naboulsi said that he would like to apologize to ArXiv and to the Stanford centre, but said that his other papers, on cosmology, were not plagiarized.

ArXiv plainly states that "authors must make their own submissions", but Ginsparg is not planning to take any action against Watanabe, who now deeply regrets the incident. "He has already suffered enough over this," says Ginsparg.

Plagiarism is extremely rare on ArXiv — only a handful of its 250,000 submissions have been withdrawn for this reason. But the incident has prompted Ginsparg to think

about how to prevent further incidents.

He says that his first step will be to take a baseline measure of similarity between the archive's documents, and so generate a warning when a paper that exceeds this threshold is uploaded. Computer software, typically developed for universities to spot student fraud, is available to do this.

"The technology is there," says Fintan Culwin, an expert in anti-plagiarism software at London's South Bank University. "The question is how much does the archive want to pay to have this service." Ginsparg is now looking for a masters or doctoral student to work on the project.

♦ www.arxiv.org

Critical comments threaten to open libel floodgate for physics archive

A paper criticizing research posted on the physics preprint server ArXiv has led experts to warn of the potential for future legal actions against the site.

In an article put on the server on 27 October, Alvaro De Rújula, a theoretical physicist at CERN, the European Centre for Particle Physics in Geneva, made a strongly worded attack on Martin Rees, Britain's astronomer royal, and his co-workers.

De Rújula alleged that Rees, who is based at the University of Cambridge, claimed credit for breakthroughs in γ -ray astronomy while ignoring contributions made by other groups. De Rújula has worked with some of those whom he says were ignored, including Arnon Dar of the Israel Institute of Technology in Haifa.

The incident raises some difficult issues for ArXiv, admit its administrators at Cornell University in Ithaca, New York. Rees was asked by Dar in September to change the citations, and says that he had originally intended to wait until

the paper had been peer-reviewed before doing so. But two days after De Rújula's posting, Rees modified his paper to cite Dar's work.

But if the criticism is seen as successful, other authors could rush to pursue similar complaints. "We don't want these food fights conducted on a regular basis," says Paul Ginsparg, the Cornell physicist who created ArXiv.

The sharp language used by De Rújula to describe Rees could cause further problems for ArXiv — which doesn't usually check papers before they are posted — if De Rújula's paper were deemed libellous in a court of law. And some experts think that it could be. "The accusations against Rees are potentially



Martin Rees: criticized online.

defamatory," says David Mascord, a media-law specialist at DMC Development, a training company based in Brockenhurst, UK.

Postings on ArXiv are deemed to be published and so are subject to libel law. In the United States, libel actions are hard to pursue because of the right to freedom of speech enshrined in the First Amendment to the constitution.

But because the paper has been published on the web, it is subject to the libel laws of any country in which it can be accessed. Under English law, for example, De Rújula, Ginsparg and Cornell University could all be liable if the De Rújula paper were held to be defamatory.

Ginsparg says that he has not taken legal advice over De Rújula's paper, but that he would remove it if recommended to do so by his lawyers. "ArXiv is just a mindless redistribution system," he adds. "It's not implemented to be a global police force to detect or enforce professional ethics."

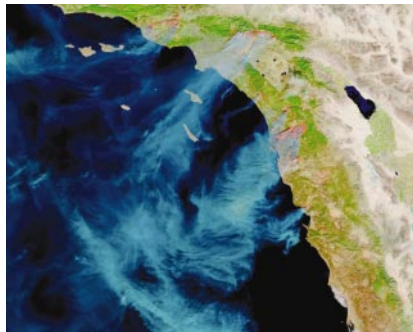
Jim Giles

California fires close labs as grant hopefuls get fingers burned

San Francisco The wildfires that ravaged southern California last week came at a bad time for San Diego researchers rushing to file their grant applications with the National Institutes of Health by the deadline of 1 November.

Five days before the applications were due, the smoke from the fires, which have destroyed about 300,000 hectares east of the city, became so dense that several major research institutions were forced to close. Among them were the University of California, San Diego, the Salk Institute for Biological Sciences and the Scripps Research Institute. All reopened within three days, but by then applicants had lost valuable time. Some researchers were also delayed by fire damage to their homes and communities.

The Salk Institute has requested a two-week extension for all of its applicants. But this now appears unlikely to be granted. The NIH stated that it would consider extensions on a case-by-case basis, but that these will not generally be longer than the duration for which the institution was closed.



No smoke without fire: wildfires have set large parts of southern California ablaze.

Climate-change defeat narrower than expected

Washington The US Senate last week defeated a climate-change bill that would have limited emissions of carbon dioxide and other heat-trapping gases. But the relative closeness of the vote encouraged many environmentalists.

Introduced by Senators Joseph Lieberman (Democrat, Connecticut) and John McCain (Republican, Arizona), the bill called for immediate cuts in emissions from public utilities and other industries.

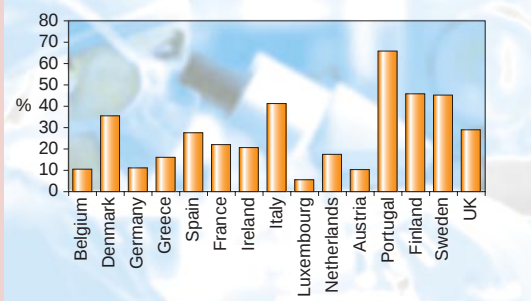
It was defeated by 55 votes to 43. But proponents of the bill, who see growing support for such emissions caps, point out that the vote is the closest that the Senate has ever come to acknowledging the problem of global warming. Six Republicans

Women struggle towards top of career ladder

Munich The European Commission has rounded up the statistics on how women are faring in science — and the news is mixed.

The report, entitled "She Figures 2003", shows that the annual growth rate for women in public research is now 8%, compared with 3.1% for men. But the fraction of women employed as academic or industrial scientists only rose from 33% to 35% between 1999 and 2001.

The number of women who participate on senior scientific panels — which closely reflects the number in top jobs — varies radically from country to country (see chart). "Normally, the higher the prestige connected with the position, the lower the quota of women," says Rossella Palomba, one of the report's authors and a



The proportion of women on senior scientific panels in 2001.

gender expert at the Institute for Population Research in Rome. In Portugal, the only country to break the 50% barrier, the authors note that the data come from a sample of just 15 positions. "This calls for an urgent review of recruitment strategies and appointment procedures," says Philippe Busquin, the European research commissioner.

crossed party lines to vote against the Bush administration's policy, which opposes mandatory controls on greenhouse-gas emissions.

Oxford don suspended for political gaffe

London A pathologist at the University of Oxford has been suspended for two months without pay after rejecting a prospective graduate student who had served in Israel's armed forces.

Andrew Wilkie rejected Amit Duvshani's application because of what he termed "gross human-rights abuses on the Palestinians" (see *Nature* 424, 7; 2003). His comments triggered an investigation that led to this week's reprimand by the university. "This ruling reflects that there can be no place for any form of discrimination within the University of Oxford other than on the grounds of merit," a university statement said.

Andrew Marks, president of the International Academic Friends of Israel, a New York-based non-profit group devoted to fighting academic boycotts of Israel, says that he is satisfied by the ruling. Marks says that Duvshani is now hoping to study at a university in the United States.

Money pours in for work to boost crop yield in droughts

London Proposals to help squeeze more crops from each drop of rain and river water were unveiled this week in Nairobi.

The details of 50 research projects aimed at tackling the effects of water scarcity on agriculture were announced during the launch of the Challenge Program on Water and Food — a multimillion-dollar scheme

spanning nine river basins in Africa, Asia, the Middle East and South America.

The plans include a study of how local communities around rivers such as the Niger in western Africa can enclose floodwaters and use that resource to farm fish. Another project will assess the impact of India's multibillion-dollar plans to divert floodwater from eastern rivers to central and western areas of the country by 2016.

The Consultative Group on International Agricultural Research, a Washington-based group that supports research to improve food security, aims to raise US\$120 million for the programme. It says that donors have already committed US\$60 million.

♦ www.waterforfood.org

Singapore slings cash at research centres

Sydney Singapore has officially opened its 'Biopolis' — a S\$500-million (US\$290-million) biological research centre that will eventually house more than 1,500 scientists.

The opening last week was followed by the announcement that the government will also create a Centre for Molecular Medicine. The centre will be run by the government's Agency for Science, Technology and Research (A*STAR), with the aim of translating basic research into medical applications. It will initially be housed at Biopolis when the centre opens in January 2004, but will later be moved to labs next to the National University of Singapore.

Kong Hwai Loong, executive director of A*STAR's Biomedical Research Council, says that the centre is expected to receive initial funding of at least S\$30 million. The centre's first project will be in using stem cells for regenerative medicine, he says.

This protein belongs to...

The early days of genomics were marked by concerns that wide-ranging gene patents would restrict research and medical discovery. So far, proteomics hasn't toiled under the same cloud. But don't get complacent, warns David Cyranoski.

A simple claim of ownership rarely sparks a wave of widespread panic. But in 1991, the US National Institutes of Health (NIH) managed just that when it claimed intellectual-property rights on some 3,500 genes, based on sequences of tiny fragments of their DNA. The news pushed the NIH researcher who had obtained the sequences, Craig Venter, into the limelight — a position that this pioneer of genomics has delighted in ever since. And it stirred fears around the world that the scientific and medical advances promised by the human genome sequence would be restricted by overarching patent claims.

So in December 2001, when the British biotechnology company Oxford Glyco-Sciences (OGS) announced that it was trying to patent more than 4,000 proteins linked to disease, you might have expected more howls of outrage. In the event, the news generated barely a ripple of interest — most probably because protein patents have been around for a century or so, without causing anyone a major headache. “We are used to them,” says Richard Gold, director of the Centre for Intellectual Property Policy at McGill University in Montreal, Canada. Indeed, patents are generally thought to provide an appropriate financial reward for those who devise useful applications for proteins, while stimulating further research.

But some experts warn that this happy balance might soon be disturbed. Massive projects are promising to solve the structures of thousands of proteins in record time. And recent rulings in US courts, where intellectual-property trends often begin, have

set precedents that could make protein patents more obstructive in the proteomic era than they have been in the past. “These trends bode ill for the future of biomedical research,” says David Korn, senior vice-president for biomedical and health-sciences research at the Association of American Medical Colleges in Washington DC.

Although they were ultimately denied, the NIH's 1991 gene patents caused alarm because of their huge potential reach. They could have meant that anyone who developed a useful product, such as a drug, by studying a particular gene would have to pay royalties to the researcher who first sequenced a tiny fragment of that gene's DNA. This, many experts argued, would provide a major disincentive to investment in research and development — the exact opposite of what the patent system is supposed to achieve.

Clamp-down on claims

Since then, the US patent office has raised the bar on gene patents. Guidelines issued in December 1999 made the condition of ‘utility’ a tougher nail to hit. Those hoping to claim a patent on a DNA sequence because it might be useful as a molecular probe or because it can make a protein now have to answer some specific questions. A probe for what, exactly? A protein that does what? “You can't just say you found a gene that might have some value,” says Tim Caulfield, an expert in healthcare law at the University of Alberta in Edmonton, Canada.

Decades of experience with patents on proteins have suggested that they don't cast such a long shadow over research and inno-



vation. And so far, at least, the fate of OGS has reinforced the general view that its aggressive patent announcement was nothing to worry about. Less than a year after claiming rights to 4,000 proteins, identified by comparing diseased with healthy tissues, OGS cut one-fifth of its staff. This spring, the cash-strapped company was bought out by Celltech of Slough, west of London. Celltech is retaining OGS's work on cancer, but is trying to sell off the rest of its proteomics operation.

Indeed, protein prospecting has turned out to be a very tough business. Large Scale Biology, based in Vacaville, California, similarly hoped to make a fortune by identifying and patenting a large number of medically important proteins. But it is now concentrating on the narrower goal of producing animal proteins in genetically engineered plants. “When you do the figures, it's just not worth it,” says Tom Gallegos, the company's senior director of intellectual property.

It costs a couple of hundred thousand dollars to patent something worldwide. “It's easy to find a lot of proteins and get patents, but with costs like that, you have to be pretty sure that your protein has commercial value,” says Yoshiji Fujita, who heads Tokyo Medical University's new Clinical Proteome Center.

Such value isn't easy to come by. Most proteins are initially patented as diagnostic markers to identify patients suffering from a particular disease or from a drug's side effect. “But diagnostic markers often don't even make enough money to pay for the



Production line: Japan's Genomic Sciences Center in Yokohama aims to churn out the structures of 3,000 proteins over the next few years.



patent," says Gallegos. "It's a lean market."

The economics change if your protein is a drug target or can itself be used as a drug. The protein erythropoietin, or EPO, for instance, is marketed as a treatment for anaemia and earns billions of dollars a year for Amgen of Thousand Oaks, California, which holds the patent on it. But to figure out whether a protein has therapeutic value generally requires extensive research in the Petri dish and in live animals. "You need hard-working people injecting things into mice," says Gallegos.

ID parade

Massive projects now under way could make identifying therapeutically important proteins much easier. The factory-like approach of Japan's 'Protein 3000' programme and the US Protein Structure Initiative should rapidly determine the structure of thousands of proteins. In theory, this explosion of structural data should help researchers home in on candidates for drug development. In August, the Japanese project, based at the RIKEN Genomic Sciences Center in Yokohama, reported that it had cranked out structures of 613 proteins in its first 18 months.

Structural data alone are not sufficient to claim a patent, however, thanks to an agreement reached in November 2002 by representatives of patent offices from Japan, the European Union and the United States. This has eased fears about overarching protein-patent claims. But projects such as Protein 3000 could still change the intellectual-prop-

erty landscape. The data from the Japanese programme, for example, will be made available through partnerships to companies in Japan before they are released internationally.

"The fear is that some proteins that have great importance in terms of research on potential therapies will be patented," says Arti Rai, an intellectual-property expert at Duke University in Durham, North Carolina. A company might then attempt to claim ownership of any approach to knocking out the protein, she says.

Some recent court rulings have started to raise concerns that patents on proteins are being interpreted too broadly. In January, Transkaryotic Therapies of Cambridge, Massachusetts, failed to break Amgen's grip on EPO when the US Court of Appeals for the Federal Circuit decided that its rival EPO product infringed Amgen's patents. These patents describe the production of human EPO in hamster ovary cells. Transkaryotic produces a slightly different version of EPO using genetically engineered human cells. So far, the court has broadly supported Amgen's claim that its patent covers any use of mammalian cells for the production of EPO.

Some observers are concerned by this turn of events. "The court is constructing a 'protect R&D investment' strategy without thinking through the implications," says Robert Cook-Deegan, director of the Center for Genome Ethics, Law, and Policy at Duke University. Such precedents could discourage companies from trying to make better

versions of protein-based drugs, he says.

For most academics, even broadly interpreted patents have held few fears. US patent law has allowed a 'research exemption' for work with patented tools or materials "solely for amusement or to satisfy idle curiosity, or for strictly philosophical inquiry". Traditionally, this has been interpreted as covering all academic research — a recent survey led by John Walsh at the University of Illinois at Chicago confirmed that researchers at US universities routinely fall back on this exemption without worrying about being charged patent royalties (J. P. Walsh, A. Arora and W. M. Cohen *Science* **299**, 1021; 2003).

Exempt no more

But in the light of a recent ruling from the US Supreme Court, says Cook-Deegan, it is unclear whether this 'gentleman's agreement' can continue. On 27 June, America's highest court ruled on the case of Duke University versus *Mayday* — and according to many intellectual-property experts, the Supreme Court's interpretation of the research exemption means that most academic research would now infringe any relevant patents.

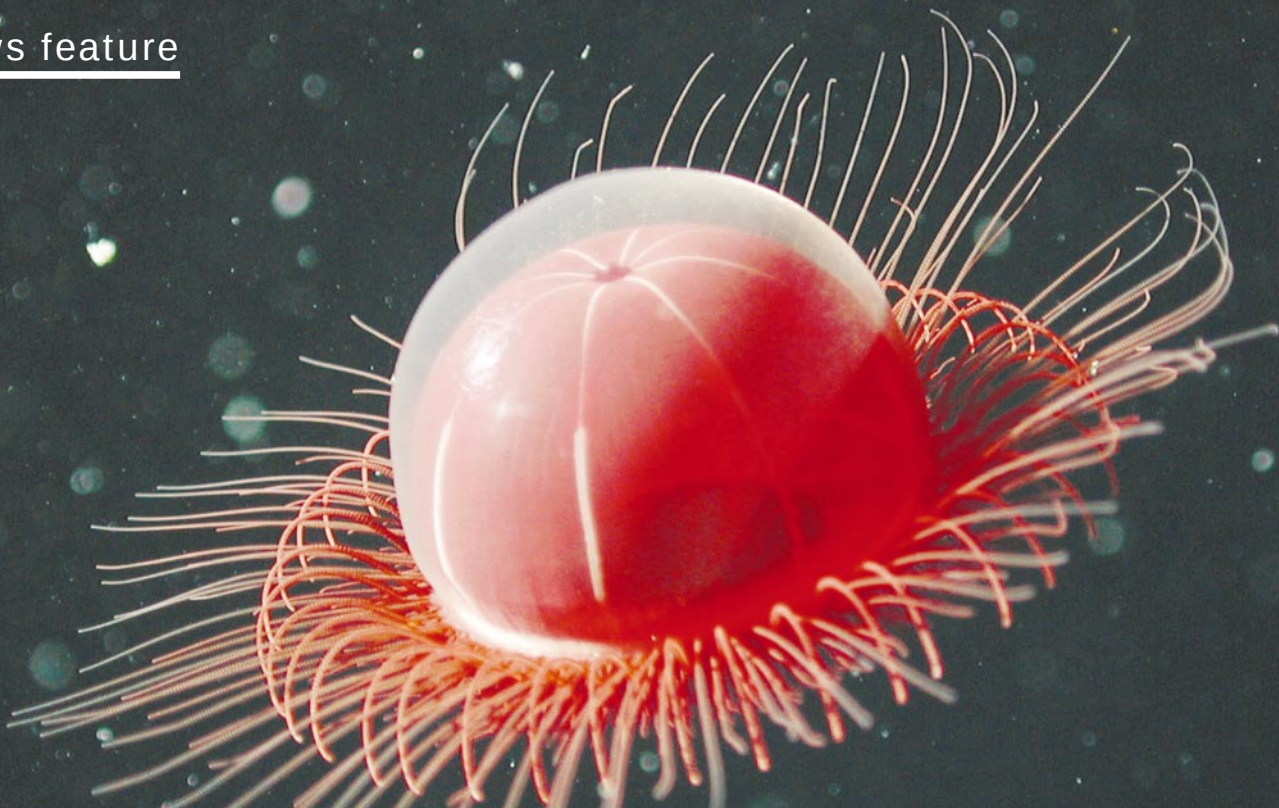
The case relates to a dispute over the use of laser technology developed by John *Mayday*, who worked at Duke until 1998, when he left for the University of Hawaii at Manoa. *Mayday* objected to Duke continuing to use the technology covered by his patents. When Duke cited the research exemption, *Mayday* countered — and the Supreme Court agreed — that the university was using the technology in the business of teaching and getting grants, not to satisfy idle curiosity.

The ruling could have onerous implications for anyone who wants to conduct research on a patented protein — and, indeed, on academic research more generally. "The case says that whether you're at the university or anywhere else, research is not just playing around. Research in and of itself is infringement," says Masashi Miyano, who heads the RIKEN Structural Biophysics Laboratory in Harima, Japan. At the very least, says Cook-Deegan, it may require extensive legal groundwork before researchers are given the green light to work on a patented protein.

The NIH is also worried about the implications of broadly interpreted patents on proteins and other biological molecules. In September, it awarded a \$1.2-million grant to Steve Merrill of the National Academy of Sciences in Washington DC for an 18-month study of the issue.

Merrill's team will consider various recommendations to policy-makers, including a stronger research exemption and the encouragement of toned-down licensing agreements that give users the freedom to use patented molecules as they wish. Policies will have to be set soon, Merrill says: "The longer we wait, the more doors will become closed to us."

David Cyranoski is *Nature's* Asian-Pacific correspondent.



Close encounters of the jelly kind

Deep-diving submarines have opened a new window onto the ocean's gelatinous inhabitants. And biologists are discovering that these denizens of the deep have a few surprises in store. Carina Dennis dives in.

Spending hours in a damp, cold and cramped space, peering through a tiny window into murky darkness, doesn't sound like a recipe for job satisfaction. But for Dhugal Lindsay, a biologist at the Japan Marine Science and Technology Center in Yokosuka, this is the best part of his job.

The reason Lindsay subjects himself to long hours in such conditions, crammed into a three-person submarine thousands of metres below the ocean waves, is because it's the best way to spy on jellyfish and other gelatinous creatures — collectively known as jellies. The ocean is awash with these diaphanous creatures, from the familiar jellyfish — properly referred to as cnidarians — to ctenophore comb jellyfish with waving cilia that propel them along, to a mass of other creatures including gelatinous snails, worms and larvaceans — tiny tadpole-like filter-feeders that spin mucous webs to catch their food. But before the advent of submersible vehicles, researchers were in the dark about this marine menagerie, and what its members get up to in the murky depths.

Only now are researchers such as Lindsay, with their undersea craft, starting to fill in the gaps. And they are startled by what they have found. Jellies, it seems, have a much bigger role in the ocean system than we thought, and could even provide the missing piece in

a long-standing puzzle about how carbon cycles from the ocean's upper layers to its floor. "We greatly underestimated their ecological significance," says Bruce Robison, who dives for jellies in waters near the Monterey Bay Aquarium Research Institute in Moss Landing, California.

Sting in the tail

For most of us, jellyfish are either simple curiosities to be admired at the aquarium, or pests that keep us from bathing in the sea. To others, they pose more of a problem. Mass blooms of jellyfish, which appear when the creatures reproduce in the spring or summer, are a perennial nuisance to fishermen and coastal dwellers. They decimate fisheries by munching on fish larvae and eating all of the shrimp and plankton — the same food that fish rely on. They stow away in ships' ballast waters and take over ecosystems in foreign seas. They can even close down coastal nuclear power plants by clogging up the pipes that bring cooling water from the sea.

Despite all this, jellies have tended to slide off the radar of most oceanographers, whose perception of mid-ocean biology has largely been based on what they haul up in nets. That hasn't included many jellies, as they are often chopped into pieces of gelatinous gunk by the netting. With no easy way to collect them,

jellies have literally fallen through the gaps of oceanographic study.

Only with the widespread introduction of research submarines in the 1980s did researchers begin to spy on the world of jellies — and scientists have still barely scratched the surface of this gelatinous underworld. Most research subs are limited to depths shallower than 1,000 metres, leaving the vast world below unexplored — the average ocean depth is about 4,000 metres and the Pacific Ocean's Mariana Trench extends down to 11,000 metres. Even the jellies above 1,000 metres have been largely neglected by submersibles. "Researchers are usually in a rush to get the subs down to the bottom as fast as possible and in the dark to save battery life, so they miss the show," says Claudia Mills, a jelly biologist at the University of Washington's marine laboratory in Friday Harbor. Mills knows her stuff when it comes to jellies — she just had a new species of deep-sea jellyfish, *Crossota millsae*, named after her¹.

Lindsay is among the few lucky biologists to have access to one of the world's deepest-diving subs — the *Shinkai 6500*, which can dive to 6,500 metres. It's a strange kind of luck, though — few people would covet the conditions Lindsay faces in this craft. A dive lasts roughly eight hours, during which the scientist, pilot and co-pilot are crammed into

a space of just 4.5 cubic metres. "It's cold and clammy, the temperature is freezing and you can't stretch out your legs," says Lindsay. Temperatures drop to less than 2 °C below 2,000 metres, so the voyagers have to wear special suits to stave off hypothermia. The vessel is not heated, both to save on battery power and to prevent the risk to its flammable gas tanks. Even looking out of the window isn't easy. There are three viewing ports, but each is smaller than a grapefruit — any larger and they would collapse under the pressure.

Juggling act

As if that isn't enough, Lindsay goes to rather extreme measures to squeeze as much useful time into his dives as possible — he counts himself lucky if he is aboard more than one of the 60 dives that the *Shinkai 6500* makes each year, such is the competition for the vessel's time. Lindsay's time-saving measures include tricks to avoid the call of nature. "Everyone has their own strategy, but I self-medicate with vodka the night before to make sure I'm dehydrated," says Lindsay. Coffee is kept to a minimum — just enough to keep the pilot alert. And Lindsay usually finds it too time-consuming and awkward to eat a proper lunch, so he sucks candy to keep his blood sugar up. He sits for hour after dehydrated hour with his forehead pressed against a window, calling directions to the pilot, manoeuvring the robotic arms that collect specimens, and trying to capture the most interesting creatures on video. "It's a juggling act," he says. Steering a 26,000-kilogram vehicle around a fragile jelly is not easy — often, jellies splatter against the sides of the craft or hurtle into oblivion when the submersible uses its thrusters. "It's always the one you really want that you can never get," Lindsay laments.

After catching the jellies, there is still the challenge of getting them out of the deep and into the light of day. "We manage to trap them in containers at these great depths, but we



Depth charge: *Shinkai 6500* gives researchers a rare glimpse of the deep ocean's delicate jellies.



rarely get anything back intact — they tend to completely disintegrate," says Richard Harbison, who studies jellies at the Woods Hole Oceanographic Institution in Massachusetts. The fragile jellies rarely survive the temperature changes and physical buffeting they experience during the trip to the surface. There are other problems too. Many jellies eat prey that glow, and to avoid becoming a beacon to their own predators, they camouflage their guts with red or black pigment. Lindsay thinks that these pigments can become lethally toxic on

exposure to light — including the headlamps on submersibles. "I had been trying to catch a particular red jellyfish for four years," he recounts. "Finally, I caught one. I brought it to the surface in a protective case and was just about to start filming when the light on the video camera inadvertently switched on at full strength. The jellyfish spat out its gut and disintegrated within minutes before my eyes."

There has to be a big pay-off for this kind of suffering and frustration, and for Lindsay it is the joy of discovery. "I see a new species almost every time I dive," he says. Discoveries can also be made after the dives — Lindsay's videos, along with footage taken by colleagues who were persuaded to turn on the sub's lights during their descent, have revealed yet more creatures of the deep. So far the tally stands at about 2,000 species of jelly — doubtless, many more lurk undiscovered in the deep.

Aside from the new species, the sheer number of jellies has caught researchers by surprise. "There are no published records on how much of the midwater biomass is gelatinous. Studies with nets suggested that it might be 1 or 2% but now, with submersible dives, it seems like it's more than 50%," says Lindsay.

Jelly fishing

Richard Brodeur, from the National Marine Fisheries Service in Newport, Oregon, was originally interested in studying fish, but ended up studying jellyfish by default. "We couldn't complete our research surveys of fish because there were so many jellyfish in the water. They'd fill our nets before we could catch the fish we were after," he says. Together with his postdoc Cynthia Suchman, Brodeur has calculated the proportion of total biomass in waters off the coast of Oregon that is accounted for by the huge jellyfish *Chrysaora fuscescens*. It turns out to be comparable to that of copepods — crustaceans that are the most ubiquitous sea animals and are thought to represent a large chunk of the ocean's carbon. "Jellies are major players in the ocean's carbon biomass," concludes Robison.

Just as revealing have been the insights that such dives bring into jellies' behaviour. Jellies have been perceived as carnivores that drift passively around, snaring prey when they chance upon it. But we now know that some jellies come up from the deep under the cover of night to gorge on plankton and krill in surface waters, before sinking back down to digest their meal at depth. Other jellies have evolved specific features to prey on their gelatinous cousins — some jellies of the genus *Beroë*, for example, are made of nothing more than a mouth-like sac with tooth-like macrocilia that chomp exclusively on other jellies. "A substantial percentage of the oceanic biomass is tied up in the bodies of jellies that are feeding on each other," says Robison, who is trying to work out who is eating whom in this tangle. "We haven't yet worked out how it ties back

D. LINDSAY

JAMSTEC

S. HADDOCK

Back home: some crustaceans, such as this amphipod, settle down and raise their young atop a jelly.

into the mainstream food web," he adds.

A more fundamental complication is that jellies seem to be involved in the direct uptake of nutrients at the bottom of the food chain. Some studies suggest that jellies can suck nutrients floating loose in the water straight through their 'skin' as a nutritious snack^{2,3}. Lindsay and his postdoc Hiroshi Miyake are currently looking at how these carbon sources affect the growth of different jellies.

Jellies' diverse roles in the food web, coupled with their immense numbers, could answer some outstanding questions about how carbon cycles through marine ecosystems. Oceanographers have long struggled to understand how organisms living on the sea floor get enough organic material to support their growth. The most obvious source is 'marine snow' — a constant deluge of tiny plankton, fish faeces, moulted exoskeletons and dead organisms that drifts down onto the sea bed. But add it up and you wind up short — there doesn't seem to be enough snow to feed all the creatures that live down there. Some estimates have pegged this shortfall at around 55%, although it is thought to vary widely⁴.

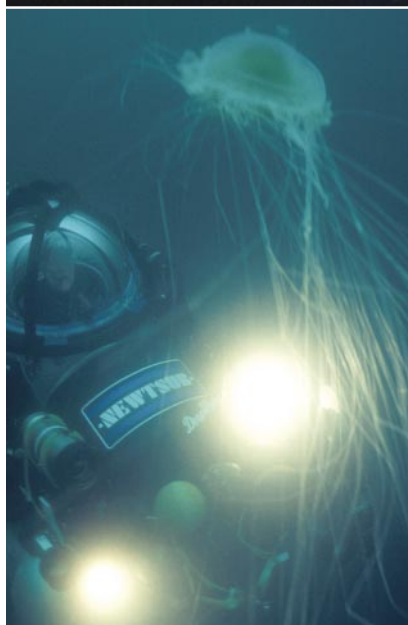
Feeding the masses

"Jellies may be an overlooked part of the equation," says Steve Haddock, an oceanographer at the Monterey Bay Aquarium Research Institute. The sheer mass of jellies, along with their ability to suck up carbon, means that dead jellies should contribute a large amount of carbon to the sea floor. Living jellies also lose globs of mucus from their exteriors from time to time, just as a human sheds dead skin. And larvacean jellies contribute to this rain of gelatinous gunk by discarding their mucous webs, which they use to catch food, when they become clogged. As these globs fall through the water column they are compressed by the increasing pressure and accelerate downwards. "These become great big, carbon-rich, fast-moving particles the size of your fist that shoot to the deep sea floor," says Robison. "The carbon content of these particles is a major percentage of what gets down there."

A final surprise has been the close relationship that jellies seem to have with other organisms. From net hauls, it was known that some crustaceans are frequently found in close proximity to each other, but it was a mystery how they manage to stay together — until observations during dives showed that they perch together on top of jellies.

It makes sense. Jellies can provide shelter and food for a huge range of creatures, who hide in the jellies' folds or even nibble on the jellies themselves. On a recent voyage to the Gulf of California, Haddock and his colleague Rebeca Gasca at ECOSUR, a marine-research college in Chetumal, Mexico, discovered that a crustacean called *Oxycephalus* nurses her young atop a comb jellyfish. And many deep-sea explorers have observed that *Deepstaria*

With the advent of diving missions, researchers such as Bruce Robison (below) are uncovering the secret watery world of jellies — some, for instance (below right), dine on jelly themselves.



enigmatica, one of the largest midwater jellyfish, which looks like a thin sac of white jelly, almost always has one or two crustaceans called *Anuropus* living with it. "We have no idea why they have this relationship or how they manage to find each other in the vast ocean," says Haddock. "There is a lot of weird stuff going on down there that we're only beginning to figure out."

To solve the remaining mysteries, jelly researchers will need more and better subs optimized for studies at intermediate depths. "Most submersibles are designed for work on the sea floor," says Lindsay. Their viewing ports often face downwards, and many of the vehicles cannot neutralize their buoyancy, meaning that they have to use their thrusters to follow an animal.

The US National Academies in Washington DC is currently assessing the future needs

of deep-sea science and the role of vehicles in future ocean exploration⁵. The report is expected to be published shortly and, although committee members have declined to comment on its contents, it is likely to include recommendations for new vehicles, or modifications to old ones, that will better suit the midwater realm. "We're poised for a golden age in studying gelatinous animals. All of the techniques are coming together and now we can start to give jellies the attention they deserve," enthuses Haddock.

In the future, much of this work will probably be done through video cameras attached to robotic submersibles. These are cheaper to build, simpler, smaller and can dive for longer, so they can collect more data than crewed missions⁶. But researchers are reluctant to stop using crewed submersibles altogether. "There is no substitute for the human eye," says Lindsay. Cameras cannot always convey a sense of depth of field, or focus on several objects at once, he points out. Despite the cramped legs, the late-night vodka sessions and the freezing cold, Lindsay would rather keep going down there himself.

Carina Dennis is Nature's Australasian correspondent.

1. Thuesen, E. V. *Zootaxa* **309**, 1–12 (2003).
2. Shick, J. M. *Biol. Bull.* **148**, 117–140 (1975).
3. Haddock, S. H. D., Rivers, T. J. & Robison, B. H. *Proc. Natl Acad. Sci. USA* **98**, 11148–11151 (2001).
4. Smith, K. L. & Kaufmann, R. S. *Science* **284**, 1174–1177 (1999).
5. <http://dels.nas.edu/deepsurgence>
6. Clarke, T. *Nature* **421**, 468–470 (2003).

LEFT: S. HADDOCK; ABOVE, D. LINDSAY
K. EVANS/NATL. GEOG. SOC.

Leonardo knew the fluid boundaries of science

Without the aid of mathematical models, his ideas stemmed from keen observation.

Sir — The Science in Culture article “Leonardo lifts off” (*Nature* **421**, 792; 2003) provided a critique of the myth that Leonardo da Vinci was “a man ahead of his time”, by suggesting that his plans for a successful flying machine design depended more on luck than his knowledge of fluid dynamics. In particular, it was stated that Leonardo “did not pay attention to the fact that air, unlike water, is compressible, and had not considered such a possibility”.

I believe that a number of writings by Leonardo indicate the contrary. On a page now in the *Codice Atlantico* (at the Ambrosiana Library in Milan), Leonardo wrote: “Water is incompressible. ... The opposite is the case for air, which when forced into vases with small openings, containing some water, ... drives away the water with such fury (*furore*) that it will be sprayed a great distance away, till that air that remains in the vase recover its initial density” (the translation from the Italian original is mine).

In the same collection there are notes made by Leonardo about the compression



Data stream: Leonardo's *Study of Flowing Water*.

of air below birds' wings during flight.

I do not think, however, that the correctness of observations such as these makes Leonardo any closer to contemporary science. Leonardo, unlike Galileo and Newton, did not rely on abstract mathematical models, capable of producing quantitative predictions, which could then be confirmed by experiment. His own studies in fluid dynamics show this very clearly. They were firmly rooted

in extremely penetrating visual observation, as witnessed by his drawings of water streams (see figure), but they totally lacked the mathematical approach of Bernoulli's *Hydrodynamica* (1738).

In Leonardo's water studies, the separation between art and science itself becomes a fluid boundary. His *Study of Flowing Water*, made in about 1509–1511, is a masterly representation of water flow carried out with supreme keenness of observation and attention to detail, powerfully conveying the sense of movement of the fluid. Nature here is not yet understood in terms of

abstract concepts and mathematical models; rather, its phenomena (and movement in particular) are directly grasped by the mind through the eye, in a manner which is closer to our experience of aesthetic contemplation than to the approach of modern science.

Stefano E. Grillo

Institute of Materials Science and Process Engineering (CNRS), Tecnosud, Université de Perpignan, 66100 Perpignan, France

Open access: other ways

Sir — John Ewing, in Correspondence, argues that open-access journals are not open to everyone because not all authors can pay or find a sponsor to pay their processing fees (*Nature* **425**, 559; 2003). Although publishers of open-access journals such as the Public Library of Science (PloS) say that authors who can't pay won't have to, Ewing feels they have underestimated the numbers who will not be able to pay.

As a long-time advocate of open access to science, I can list several points to suggest that this concern is overstated.

Declan Butler notes in his News Feature “Scientific publishing: who will pay for open access?” (*Nature* **425**, 554–555; 2003) that many funding organizations are willing to pay these fees for their grant recipients. Although this solution will not work in disciplines that are less well-funded than medicine (in Ewing's field of mathematics, for example, or mine, philosophy), that is no objection to applying it to fields where it can work. It is highly likely, and desirable, that different fields will develop different open-access models. Some peer-reviewed open-access journals in the humanities charge no processing fees at all.

Ewing notes that many universities may not be able to afford these fees, in cases

when funding bodies cannot pay them. But if open access spreads, every university will make large savings from the conversion, cancellation or demise of expensive subscription-based journals. This money can be used to support the open-access model of archiving and publication.

If some future open-access publishers have no policy to waive fees, and authors find themselves excluded on financial grounds alone, there are other ways to bring about open access to peer-reviewed literature (such as e-print archiving) that do not depend on open-access journals, research grants, affluent employers or windfall savings in the library budget.

Peter Suber

Department of Philosophy, Earlham College, Richmond, Indiana 47374-4095, USA; Public Knowledge, 1875 Connecticut Avenue, NW, Washington DC 20009, USA

Open access will deter illegal file-sharing

Sir — John Ewing argues in Correspondence (*Nature* **425**, 559; 2003) that, in the current reader-pays publishing system, researchers without a journal subscription can obtain an article by other means. This is probably

true. However, article-sharing is an illegal and unreliable method of getting scientific information, because many commercial publishers own copyright as well as the rights to distribute the results. The magnitude of this copying activity is, as far as I know, unknown.

A few requests to “send me a PDF file of an article” will probably not hurt the current system. However, if this approach were to become more systematic, a PDF file (or other formats) could be ‘shared’ using databases and peer-to-peer networks just as any MP3 music file can be shared worldwide, to the great annoyance of the music industry. This would definitely hurt the current subscriber-pays system.

The open-access alternative is immune to such copying activities because the articles are available free of charge.

Martin Dufva

MIC-Mikroelektronik Centret, Oersteds Plads, Technical University of Denmark, Building 345 East, DK-2800 Kgs. Lyngby, Denmark

The Nature Publishing Group does not retain copyright. Authors grant NPG an exclusive licence; in return they are free to reuse their papers in their future printed work, and to post a copy on their own not-for-profit website. See <http://npg.nature.com/authornews> — Editor, *Nature*.

The history of cool

A look back at pioneering studies of quantum effects at low temperatures.

Cold Wars: A History of Superconductivity

by Jean Matricon & Georges Waysand
Rutgers University Press: 2003. 304 pp. \$65
Philip W. Anderson

This book, translated from the French, is a history of the field of cryogenics and of the two specifically quantum forms of matter in bulk — superconductivity and superfluidity (although only the first is mentioned in the subtitle). The authors clearly express their enthusiasm (which I share) for the field and its history: the fascination and mystery that it held for the icons of the quantum revolution, from Albert Einstein through Werner Heisenberg to Richard Feynman; the slow and uneven advance of science as it was impeded by personality conflicts, the savagery of the Stalinist regime, and the politics of the real Cold War; the clashes of conflicting ways of doing research in disparate fields; and even, most recently, a case of the deliberate manufacture of scientific results on an unprecedented scale. There is a lot of rich material here that has been, on the whole, neglected.

This history has more than its share of fascinating, larger-than-life figures. It begins with the Dutch entrepreneur Heike Kamerlingh Onnes, who discovered superconductivity. In order to liquefy helium, he created in Leiden the Cryogenic Laboratory, the first industrial-scale lab for pure scientific research and the true precursor of today's CERN, Fermilab and Kamiokande.

Then there were the Russians: Pyotr Kapitsa, Ernest Rutherford's favourite, who was kidnapped by Stalin from Cambridge, UK, to carry on his helium research in Russia; Lev Shubnikov, who founded the great Kharkov lab but was murdered by Stalin in 1938, some 15 years before his wonderful experiments demonstrating 'type II' superconductivity in alloys — the phase that makes possible high-field magnets such as those used in magnetic resonance imaging (MRI) — were understood; and the brilliant figure of Lev Landau, Russia's greatest theorist.

We also meet Fritz London, the under-recognized genius who was denied recognition for his great contribution to superfluidity — the realization that it was a Bose–Einstein condensate — by Landau. There is John Bardeen, who, with Leon Cooper and Robert Schrieffer, proposed the BCS theory of superconductivity, but only after first winning a Nobel prize for the transistor. And finally we come across the charismatic Bernd Matthias, the 'alchemist' guru of what the authors call the "age of materials". Matthias



Super stars? Researchers at Heike Kamerlingh Onnes' lab in Leiden discovered superconductivity.

died too early to catch the breakthroughs in high-temperature superconductivity made by his followers and students.

Having said all that, throughout the book I kept encountering historical blunders and misapprehensions that left me wondering how sound the rest of it was. For instance, in the space of two pages there are three dubious historical judgements. To begin with, "The history of physics was indelibly marked by WWII, primarily because of the ... Bomb," write the authors, but more physicists worked on radar than on the bomb, and much more technology resulted from it. "Los Alamos made no contribution to low-temperature physics," the book says, but research on helium-3 originated at Los Alamos, and Matthias' laboratory there played an important role later. And the authors declare that, in contrast to the experimentalists, "it was essentially the great pre-war figures who continued to hold center stage among the theorists." But the immediate post-war generation of theorists was arguably as strong and as numerous as any in history: Feynman, Tsung-Dao Lee, Chen Ning Yang, Julian Schwinger, Murray Gell-Mann and, among the condensed-matter types, David Pines, Philippe Nozieres and Walter Kohn.

There are also errors in personal details that in many cases subtly alter the emphases in the story. Bardeen was not a "student of Slater"; he was a junior fellow at Harvard, and no other source has him seriously influenced by John Slater. Ted Geballe was not an "early student" of Matthias; rather, Geballe was his department head and mentor at Bell Labs, and the relationship was two-way. (Geballe was, incidentally, trained at the

University of California, Berkeley, a laboratory that the authors condemn to 'also-ran' status in a parenthetical note.) And the authors ignore Matthias' continuous connection to Bell Labs from before he began studying superconductivity.

The book also provides a false impression of the discovery of the Josephson effect. Brian Josephson's paper predicting tunnelling supercurrents was not ten years after BCS theory — a fact that is given some significance — but five. He was also unable to put his own predictions into experimental practice, and had a public argument as to their validity with Bardeen in the summer of 1962; this was shortly settled by experiments done by John Rowell and me at Bell Labs.

But Josephson makes a late arrival (in chapter 18 of 22) in *Cold Wars*, which focuses mainly on the period before and immediately after the Second World War. It is almost worth reading the book just for the distressing story of London, whose books laid out the problems that the post-war generations were to solve but who died just before the solutions were to become clear. There is also an excellent view of the prickly personality of Landau and of the terrible dangers that threatened him, Kapitsa and Shubnikov as they laboured well in advance of their Western colleagues from 1935 to 1955.

In describing the controversies about the theory of liquid helium's superfluid phase, the authors include the contributions of Feynman but not the penetrating insights from the early 1950s of Lars Onsager, which in my opinion are of equal status. The story of Bardeen's "relentless pursuit" of the solution to superconductivity is well told,

as is the description of the nature of the BCS theory (with a debt here to Victor Weisskopf, whose explanation is quoted).

The story thereafter becomes sketchy indeed, and misses many vital points. I might suggest that the authors' relative unfamiliarity with the anglophone world, and their weakness in theory, begin here to warp the coverage. There is emphasis on Pierre-Gilles de Gennes' group in France, with its remarkable collective ethos and a significant number of detailed applications of the BCS ideas to its credit, but does this work stand out so much relative to many things that at the time seemed more important? And I cannot let pass the authors' failure to note that although Alex Müller's great discovery of high-temperature superconductivity in the cuprates was unquestionably motivated by bipolaron theory (not an original concept of Benoy Chakraverty, by the way), that theory is nonetheless generally thought to be wrong. This is far from the first time since Christopher Columbus that a wrong concept motivated a great discovery.

In the discussion of the state of theory in this field, my words in a 2001 article for a Nobel symposium are quoted out of context, misreading or misunderstanding the message that the article was meant to convey, namely, that the source of high-temperature superconductivity is not a mystery, and that theory has not been pointless and futile. The reasons why the misleading popular impression of chaos and controversy in this field is so hard to dispel is not explored here, which is a pity. Still, in these final chapters the authors make some telling points about the overselling of the hopes for practical applications that characterized this period. But this critical observation is never balanced against the value of the MRI industry and other applications of superconductivity — by no means a unique story in the history of tensions between research and technology, where hitting the jackpot is a rarity but adds enormous value when it happens.

This year's Nobels bring out both the strengths and the weaknesses of the book. On the one hand, it is an excellent source for the background of the physics prizes to Vitaly Ginzburg and Alexei Abrikosov. But the intricate history of helium-3, now the source of yet a second physics Nobel, to Tony Leggett, is barely mentioned. I have already noted the absence of MRI, the subject of the physiology prize, from their horizon.

The intriguing piece of scientific history in *Cold Wars* has not been as well presented elsewhere, and the book is worth the attention of layman as well as scientist. But *caveat emptor*: the real inside story is not here if you're interested in what actually happened or in just who did what. But I found it refreshing to find judgements as to the broad trends of socio-scientific history, even if some of these were off by a little or, occasion-

ally, a lot. Far too often the history of science confines itself to bare facts — when it pays attention to them at all.

Philip W. Anderson is in the Department of Physics, Princeton University, Princeton, New Jersey 08544-0708, USA.

The rise and fall of populations

Complex Population Dynamics: A Theoretical/Empirical Synthesis by Peter Turchin

Princeton University Press: 2003. 456 pp. \$75, £52 (hbk); \$29.95, £19.95 (pbk)

Nils Chr. Stenseth

People have been fascinated and puzzled for centuries by the profound variations from one year to the next in the abundance of lemmings and populations of hares and lynxes. The archbishop of Uppsala, for example, wrote about the phenomenon as long ago as the fifteenth century. And hunters and other rural people such as the Sami of northern Scandinavia have their own theories to explain the burgeoning populations of lemmings, for instance, in some years. But it was the Oxford zoologist Charles Elton who built the scientific platform for the modern study of population cycles with a 1924 publication in the *British Journal of Experimental Biology*. That paper and his 1942 book *Voles, Mice and Lemmings* have been key references ever since.

Much of this work has focused on vertebrates of the north, but similar phenomena have been observed for species in other regions, such as the larch budmoth found in the Alps. The scientific literature on population cycles is vast and has to some extent been characterized by heated debates. So a book that aims to synthesize this rather chaotic field and make it more accessible to outsiders is to be welcomed.

Complex Population Dynamics, written by Russian-born ecologist Peter Turchin, is split into three parts: theory, data and finally a series of six case studies. The theory part provides an excellent synthesis of the work by Turchin and colleagues, but specialists

on population cycles may find that it does not cover the literature as fully as they might like in a book with ambitions of providing a synthesis of the field. I preferred the second part of the book, which covers both phenomenological (time series-based) and mechanistic modelling — the latter more fully than the former.

The section on case examples is good for the systems that Turchin has worked on himself, but is rather shallow for some of the other systems described, a good exception being the chapter on grouse. However, I think that this book contributes profoundly to the literature, in particular with its emphasis on integrating statistical analysis, theoretical modelling and experiments, rather than relying solely on experimental work. I fully agree with Turchin's conclusion that ecological investigations of population cycles and similar phenomena should start with statistical data analysis, aimed at describing the patterns to be explained, and end with experimental work to discriminate between alternative mechanistic explanations. In this respect the book may have a huge impact on the field, not necessarily because everybody agrees with Turchin's conclusions, but because he provides examples of what a research programme ought to look like.

Turchin's book covers many of the same elements as *Population Cycles* (Oxford University Press), which was published last year. Edited by Alan Berryman, who contributed an opening chapter and a postscript, *Population Cycles* comprises seven chapters written by specialists in the field, each considering an example of a population cycle. These chapters, although somewhat variable both in form and quality, display great enthusiasm in attempting to understand why some species and populations exhibit extensive population cycles whereas others do not. A concluding chapter is written by ecologists Xavier Lambin, Charley Krebs, Robert Moss and Nigel Yoccoz, who favour the experimental approach over statistical data analysis. Together with the postscript by Berryman, this provides a good balance regarding methodological approaches — a balance that I am convinced is needed if we are to find the solution to the cycle puzzle.

Both of these books show that the study of population cycles is a stimulating field, with much data and several plausible hypotheses needing to be tested by specifically designed and well-planned experiments. Although both books will provide active scientists in the field with much to think about — and to disagree with — I would not recommend either of them as textbooks, as they both seem too biased or narrow to serve a general educational purpose.

They both express rather similar perspectives, for instance in emphasizing feedback interactions between different trophic levels (between plant and herbivore, predator and



Now you see it... Many suggestions have been made to explain the lemming's disappearing act.

Exhibition

Force of nature

The photograph shown here, called *Forces #7*, is one of a series staged and shot by New York-based artist Sonja Braas. Braas is interested in the way that we perceive our natural environment, particularly those uninhabitable landscapes where danger is a counterpart to beauty. For the *Forces* series she built models depicting the raw violence of nature. Once Braas had photographed the models, she destroyed them.

Forces #7 is a compelling, ambiguous image, with a reality that becomes less clear the more closely one looks. It reflects the fading of the romantic notion that parts of nature will always remain pristine and inaccessible to humans.

Some of the *Forces* series can be seen at the Tanit Gallery in Munich, Germany, from 6 November to 13 December 2003.

Alison Abbott



prey, and parasite and host). But do such trophic interactions provide a general explanation of population cycles? In my view, much more work (involving all elements of the research programme advocated by Turchin) is needed to settle this question.

Turchin claims that ecology has become a mature science, but I think it is still maturing, and as such is all the more exciting to work in — it is during the maturing stages of any life cycle that the interesting developments happen. I am quite sure that Elton and the group around him in the Bureau of Animal Population, which he set up at Oxford University, would have agreed and been happy about the development of what they started three-quarters of a century ago. ■

Nils Chr. Stenseth is at the Centre for Ecological and Evolutionary Synthesis, Department of Biology, University of Oslo, Blindern, N-0316 Oslo, Norway.

Rambling in the rain

The Cruise of the Betsey with Rambles of a Geologist

by Hugh Miller, with a preface by T. C. Smout and an introduction and notes by M. A. Taylor
National Museums of Scotland: 2003.
576 pp. £20

Philippe Janvier

It is rather unusual to find poems, folklore and polemic in the same book as an analysis of the structure of the scales from 370-million-year-old lobe-finned fishes. But Hugh Miller's *The Cruise of the Betsey with*

Rambles of a Geologist contains both. It is rather disconcerting at first but it soon becomes fascinating once you imagine you are making your way with the author, who passes from one topic to another according to either the area he's visiting or an incidental encounter with a curio of nature.

The Cruise of the Betsey is mainly an account of Miller's geological and journalistic exploration of the Scottish isles of Eigg, Rum and Skye in 1844 and 1845; *Rambles of a Geologist* relates to voyages he made in the late 1840s through Scotland to Caithness and Orkney in search of geological and palaeontological outcrops. The two texts were first put together in 1857 by Miller's wife, Lydia Fraser, after his suicide in 1856. This new edition is enriched with a foreword by T. C. Smout, Scotland's historiographer royal, and with a detailed and informative introduction by M. A. Taylor, a palaeontologist at the National Museums of Scotland. It also contains a geological timetable, a glossary, notes and an index, although there is no simplified geological map of the areas being considered.

Miller was a self-taught man. As a young stonemason he became fascinated by the fossils he found around his native town of Cromarty in Scotland, especially the Devonian (470-million-year-old) fish and Jurassic (150 million-year-old) molluscs. In learning about geology and palaeontology he found no conflict between his Calvinist faith and the history of life told by the rocks. But by rejecting a literal reading of Genesis about the history of life on Earth, Miller helped to promote science within the strongly religious society of his time. He was a talented writer, and his books on this subject, *The*

Old Red Sandstone (1841), *Footprints of the Creator* (1849) and *Testimony of the Rocks* (1857), were successful in Victorian times.

Even so, Miller never adhered to the pre-darwinian evolutionary views of his time. To both Miller and his mentor, the palaeontologist Louis Agassiz, evolution, if it occurred at all, was merely a form of degeneration. In *The Cruise of the Betsey*, Miller rarely alludes to this; most of his scientific considerations are about geology, its bearings on our knowledge of the vastness of time, and its moral or physical benefits. Thanks to his popular style, this book has long increased people's interest in geology, hence the importance of this new edition.

Readers not from Scotland might have been lost in Miller's allusions to church and land politics without the book's excellent introduction. This explains that the Free Church of Scotland was created by the 'disruption', in which Miller played a role as the editor of the evangelical newspaper *The Witness*. The 'clearances' were a little-known (outside Scotland) form of 'soft' ethnic cleansing, which Miller condemned.

The diversity of subjects dealt with in this book is immense, including sociology, folklore, poetry, Gaelic language, archaeology, history, politics, religion, morality, zoology, geology, palaeontology and geography. This breadth makes it impossible to dissect the book into particular sections (as Taylor puts it, it is like trying to cut quicksilver). Yet Miller's style is clear and steady, with great care for detail. It variously recalls Walter Scott, Balzac and sometimes even Jules Verne. Although undoubtedly Victorian, Miller's English is remarkably easy to read for non-native anglophones.

Miller often alludes to rainy days ("another rainy day, varying only from the preceding day by the absence of wind"), so put on your boots and raincoat (or plaid) and just wander with him along the shore. Listen to him, inspired by a dead, tortured fish on the beach, considering the origin of moral evil, or, in a cave on Eigg, meditating over a sixteenth-century clan massacre, or explaining why Cromwell's helmet and Devonian fish scales were similarly fluted.

I warmly recommend this marvellously rambling book, which is full of sensitivity and poetry, to anyone who loves Scotland or is a humanist, a sociologist, an ethnologist, a geologist, a palaeontologist or just a fossil fan. ■

Philippe Janvier is at the CNRS, Département Histoire de la Terre, Muséum National d'Histoire Naturelle, 75005 Paris, France.

More on Miller

Celebrating the Life and Times of Hugh Miller

edited by Lester Borley
Cromarty Arts Trust/Elphinstone Institute,
University of Aberdeen, £13.50

Following the thread

What was your first experiment as a child?

I found some ants that were walking in a row. I wanted to see what would happen if I picked up and moved a leaf they were walking over. They went crazy and wandered all over the place. I was afraid they might bite me so I left.

Whose graduate student would you most like to have been (historical impossibility notwithstanding)?

Robert MacArthur. He was a very creative, insightful and curious man who had a passion for science. I understand that he was caring and supportive of graduate students and encouraged them to follow their passion.

What single scientific paper or talk changed your career path?

A talk about global warming by Stephen Schneider at a US Fish and Wildlife Service conference in 1990.

What book has been most influential in your scientific career?

Theoretical Ecology, a book edited in 1981 by Robert May.

What gives you the most job satisfaction now?

Helping to recognize and solve problems that have an impact on the natural world. Ecosystems provide humans with services necessary for our lives, and I have the privilege of working to try to ensure that these systems are not too badly damaged by humans.

What are your major frustrations?

Academia is a very competitive and judgemental field. Interdisciplinary work is needed to help solve the world's problems, yet universities and funding agencies continue to promote disciplinary fiefdoms.

What's your favourite conference destination, and why?

Snowmass, Colorado, in the summertime. It is in the Rocky Mountains, far enough from Aspen to be out of the crowds but close enough to take advantage of the cultural activities and restaurants.

What was the worst/most memorable comment you ever received from a referee?

After reading a draft of a paper I was getting ready to publish, a colleague said to me something like: "You obviously went to a public, not a private, school. Otherwise you would write better."

What book is currently on your bedside table?

The Forgotten Founders: Rethinking the History of the Old West by Stewart Udall.

You are on a plane behind two students obviously going to the same conference, who start to talk about your work. What do you do?

Let them talk and enjoy themselves.

The Internet is the bane of scientists' lives because...

...people can blast off e-mails that they have not completely thought through and are often unnecessary. I could cut out at least half of my e-mails and still get the information that I need from the other half.

What do you do to relax?

Travel to strange and exotic places, such as Bhutan, with my husband and friends to watch birds. On a day-to-day basis, I knit in any and all meetings and seminars.

Please tell us more about the theory, practice and etiquette of knitting during lab meetings. How many stitches/garments have you racked up during meetings in your career as a scientist?

The theory is that by knitting, which is more productive than doodling, I pay better attention to what is going on in the meeting or seminar. The practice is that I only knit items that I do not have to think about. By this I mean that the pattern I am knitting is easy to follow. If I knit something that takes concentration, then I do not follow what is going on in the meetings. As to etiquette, if it is a formal meeting, I will ask the speaker or organizer if they mind if I knit. I have had several members of audiences stop me and comment that I was being rude — doing the equivalent of reading the newspaper. Every time this has happened, I have spotted that on their notepad they have many doodles, so I ask if they could listen and doodle at the same time. As to how many projects I have done in this way, it probably works out at three or four projects a year, such as sweaters, vests, baby blankets and baby clothes. I began this habit around the year I went up for tenure.

What would you have become, if not a scientist?

I don't know. Back in 1982 when I was accepted into grad school at Princeton, I told them that I did not want to become an academic. I believe that I was one of the first few grad students who were accepted and said that they did not want to be a professor. I wanted to be a scientist working on environmental issues for a large company such as Arco or BP. But if I were not a scientist, I guess that I would have become a computer geek.

What single discovery, invention or innovation would most improve your life?

Hiring a personal chef would make all the difference in my immediate life. At another



Terry L. Root

Terry L. Root is a senior fellow in the Center for Environmental Science and Policy at the Stanford Institute for International Studies. She is an ecologist who specializes in conservation biology, biogeography and ornithology.

level, the invention of viable alternatives to our fossil-fuel addiction.

Do you have a burning ambition to do or learn something of no practical or immediate value?
See as many birds in the world as possible.

Under what conditions do you have your greatest and most inspired ideas?

Just as I am going to sleep — it causes sleepless nights.

What's the most interesting thing in your fridge?

Twelve containers of coffee yoghurt. Besides knitting, I am known for eating coffee yoghurt for lunch at work each day.

Why is physics so hard?

It isn't. Our culture has instilled in the majority of people that mathematics and science are impossible to understand.

What music would you have played at your funeral?

I would have a friend of my husband pick *Freight Train* on the guitar, the song my husband always plays each time he picks up the guitar.

What's just around the corner?

A mass extinction event regardless of what we do. We do, however, have the ability to lessen the number of species that could go extinct.

What would you have written on your grave-stone?

She helped.

Over the edge?

Len A. Fisk

The two Voyager spacecraft are heading beyond the bounds of the Solar System, and Voyager 1 may now have encountered the 'edge' — the termination shock — of the solar wind. But not everyone agrees.

There is adventure afoot, with controversy to boot. The two Voyager spacecraft were launched by NASA in 1977, and, having explored the outer planets, are now on a path heading out of the Solar System. Voyager 1 is more than 13 billion kilometres away, more than 85 times farther from the Sun than the Earth is. At that distance, argue Krimigis *et al.*¹ on page 45 of this issue, the spacecraft has encountered the so-called termination shock of the solar wind, where the solar wind begins to merge with the interstellar medium. But on page 48, McDonald *et al.*² argue equally convincingly that the termination shock still lies ahead. Either way, Voyager 1 has entered a region of our Solar System that has never yet been explored.

The outer atmosphere of the Sun expands continuously into space as a supersonic, flowing plasma of charged particles, known as the solar wind. The wind's velocity is typically between 400 and 750 kilometres per second; in comparison, the speed of sound near the Earth is only 30–50 kilometres per second. As two plasmas do not readily penetrate each other, the supersonic plasma of the solar wind carves out its own volume inside the plasma in the interstellar medium, a volume that is known as the heliosphere.

At some point in the outer heliosphere, however, the solar-wind plasma must begin to merge with the plasma of the interstellar medium. As with all supersonic flows, this merger begins with a shock transition at which the speed drops abruptly, from supersonic to subsonic — just like the shock that precedes a supersonic plane and causes a sonic boom. This is the termination shock of the solar wind (Fig. 1). It surrounds the entire heliosphere, and is by far the largest shock transition in the Solar System. Estimates³ of the distance to the termination shock from the Sun range from 85 to 120 AU (1 AU, or astronomical unit, is the mean distance between the Earth and the Sun, equivalent to around 150 million kilometres).

The termination shock is expected to be a fascinating astrophysical object and a prodigious accelerator of energetic particles. Interstellar neutral gas penetrating into the inner heliosphere becomes ionized, is picked up by the solar wind and carried into the outer heliosphere. These particles are then accelerated, with their energies increased by a factor of more than 10,000, to form an energetic particle population known as the

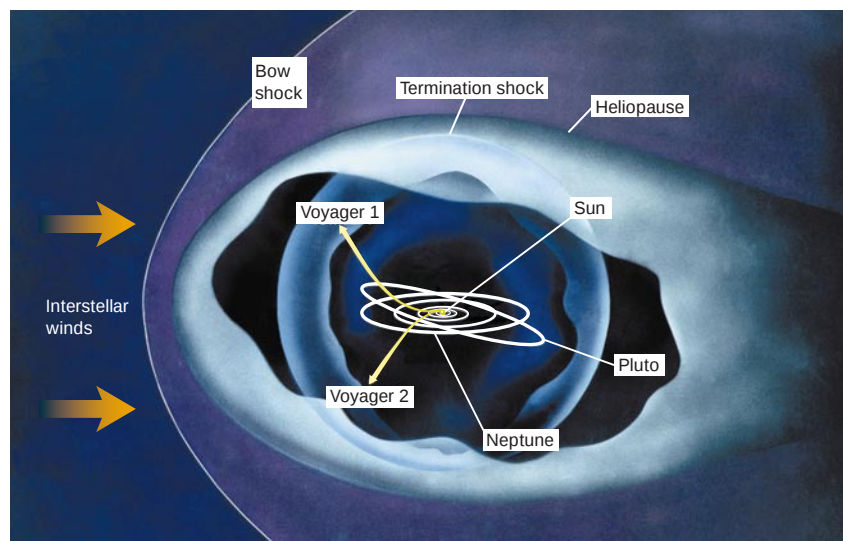


Figure 1 Fantastic voyage. Voyagers 1 and 2 have flown on different trajectories past the outer planets of the Solar System, and Voyager 1 may¹ — or may not² — have reached the termination shock, at 85 AU from the Sun. The termination shock occurs where the supersonic plasma of the solar wind begins to slow down as it encounters the interstellar medium at the bounds of the Solar System. The solar wind and the interstellar medium do not merge easily, so further out, beyond the termination shock, there is the true boundary between the solar wind and the interstellar medium — the heliopause. Further out still, if the Solar System is itself moving supersonically relative to the interstellar medium, there may be a large bow shock.

anomalous cosmic-ray component. It is likely that this acceleration occurs at the termination shock. The accelerated particles can, in turn, exert pressure that alters the structure of the termination shock. In this sense, the termination shock should be similar to shocks in the interstellar medium that result from supernovae, which are also expected to be strong accelerators of energetic particles.

If Voyager 1 has encountered the termination shock, there should be several distinct signatures in the data collected by the spacecraft. (Voyager 2 currently trails Voyager 1 by about 20 AU, and has not yet reached the possible location of the termination shock.) The most obvious signature would be the drop in the solar wind speed, with a resulting increase in density and magnetic-field strength. Also, many accelerated particles, particularly the anomalous cosmic rays, should be present.

These two different signatures are the basis for the controversy between Krimigis *et al.*¹ and McDonald *et al.*². The plasma detector on-board Voyager 1, which would have provided a direct measure of the solar wind speed, has not operated for some

years. To compensate, Krimigis *et al.* have performed a clever analysis of low-energy particles, and conclude that the solar wind speed has dropped, as required. They also see an increase in the number of these low-energy particles, and of the right types of particle, as would be expected near the termination shock. In contrast, measurements of anomalous cosmic rays at higher energies by McDonald *et al.* imply that these particles must be accelerated at a distance beyond Voyager 1's current position. McDonald *et al.* do see a substantial increase in the intensity of ions and electrons at around 85 AU, but argue that it is only a precursor of the termination shock that still lies ahead.

In fact, the termination shock is not expected to be stationary. Its location will vary with changing heliospheric conditions, which in turn respond to changing solar conditions. Indeed, Krimigis *et al.* argue that Voyager 1 only penetrated beyond the termination shock for around 200 days, after which the shock moved outwards again, leaving the spacecraft back in the supersonic solar wind. The termination shock may well be moving outwards for the next few years, in

NASA/VOYAGER

which case the chase is on again to encounter and study the largest shock transition in the Solar System.

The question is, of course, who's right? Has Voyager 1 already encountered the termination shock? Personally, I tend to agree with Krimigis *et al.*¹ that their data can most readily be explained if the termination shock had been crossed. The data of McDonald *et al.*² might then suggest that the termination shock has a more complex shape than expected, or that there is some other means of accelerating energetic particles, beyond the location of Voyager 1. Neither explanation is certain, and we must hope that Voyager 1, on its march ever outwards, will encounter this interesting region again soon.

Once the termination shock has definitely been passed, the adventure enters a new phase. The region of subsonic solar wind beyond the termination shock will be fascinating, characterized by turbulence and

particle acceleration, and by many unusual plasma phenomena. The Voyager spacecraft move at 3–4 AU per year, and will eventually encounter the heliopause (estimated to be around 150 AU from the Sun), which is the boundary that separates the solar-wind plasma from interstellar plasma (Fig. 1). Then we will have truly penetrated the interstellar medium. The Voyagers might eventually reach a 'bow shock', caused by the heliosphere itself moving supersonically through the interstellar medium. But both craft will have exhausted their power supply long before that, in around 2020. After a 40-year mission, they will drift quietly into the pristine interstellar space beyond. ■

Len A. Fisk is in the Department of Atmospheric, Oceanic, and Space Sciences, University of Michigan, Ann Arbor, Michigan 48109-2143, USA. e-mail: lafisk@umich.edu

1. Krimigis, S. M. *et al.* *Nature* 426, 45–48 (2003).
2. McDonald, F. B. *et al.* *Nature* 426, 48–50 (2003).
3. Stone, E. C. *Science* 293, 55–56 (2001).

Evolutionary biology

Scramble for the eggs

Matthew J. G. Gage

Copulating cockerels can, it seems, tailor their ejaculate to several factors: the degree of male competition, whether they have mated with the female before, and the female's reproductive 'value'.

Sperm competition — in which the sperm from more than one male compete to fertilize a female's eggs¹ — is often the final stage in the long struggle for males to reproduce their genes. It's therefore not surprising that it has favoured some remarkable male reproductive traits². One general adaptation is the production of astronomical numbers of tiny sperm, because a fundamental mechanism of sperm competition follows the scenario of a raffle: the males that ejaculate the most sperm are more likely to succeed³. But sperm production is not limitless, so males cannot perpetually achieve numerical superiority. Other tricks must be used. For instance, as Pizzari *et al.*⁴ describe on page 70 of this issue, cockerels show "unprecedented sophistication" in balancing the need to achieve numerical dominance in sperm competitions, while operating within a limited budget. At each mating, male fowl are sensitive to local social and sexual cues: they produce an ejaculate that is optimized to the level of sperm competition and the female's reproductive value, and which is balanced against their own sperm budget.

Males produce copious numbers of sperm, but the process is costly in many ways. For instance, each sperm cell is under intense evolutionary pressure to be a specialized and competitive DNA-carrying unit. Such specialization may mean protracted periods

of production: in mammals, it takes between 5 and 11 weeks (depending on the species) to manufacture a mature sperm cell from the male germ line⁵. In addition, non-trivial energetic requirements have been identified. In adders (*Vipera berus*), for example, the loss of body mass during the sperm-production stage (when males are not physically active) is as significant as that during the ensuing energetic phase of searching, courting, competing and mating⁶.

Given these constraints, evolutionary biologists have predicted that males should be prudent in producing an ejaculate, tailoring it to the specific and local demands of each mating^{7,8}. Indeed, there is experimental evidence that males can modulate sperm numbers in a seemingly adaptive manner^{9–12}. But the study by Pizzari *et al.*⁴ reveals that male fowl (*Gallus gallus*) can adjust sperm numbers over more sensitive scales and cues than has been shown before.

The red jungle fowl, and its domesticated descendant the farmyard fowl, has evolved a promiscuous mating pattern in which females mate with several males throughout their fertile window. Males can be either dominant (with full mating access to females) or subordinate (with reduced access). Such promiscuous mating patterns are common across the animal kingdom, with species that evolve true monogamy

being in the minority². For males, female promiscuity means that mating success does not guarantee fertilization success.

So how do cockerels increase their chances of fertilizing an egg when competing with rival males? To find out, Pizzari *et al.* studied free-ranging populations of feral and red jungle fowl. The birds were habituated to humans, and females wore a harness that allowed naturally produced ejaculates to be collected and analysed with minimum interference.

In one experiment, which looked at the effects of the degree of sperm competition, cockerels were allowed to copulate in the presence of none, one or three rival males, with increasing male number presenting a rising level of competition. Pizzari *et al.* observed different ejaculatory responses according to the male's social status. Dominant males — predictably — ejaculated more sperm as the risk of competition increased, presumably in an attempt to achieve numerical superiority over their rivals. Subordinate males, however, showed a more complex response: ejaculate size increased when the level of sperm competition rose from low to medium, but fell when the intensity of competition was greatest.

These responses fit closely with two different theories for predicting male ejaculatory responses. Dominant males behave as per a 'risk' model⁷, where rising risks demand increasing spermatozoal investment to win fertilization competitions. But subordinate males must play to a more sophisticated 'intensity' model⁸. Low social status means low control over females, so that high numbers of rival males predict very high probabilities of intense sperm competition. Under such scenarios, subordinate males are predicted to reduce their investment and conserve sperm for future matings, because the costs of outnumbering dominant rivals in these sperm competitions are too high. Using the raffle analogy, if there is no competition for the prize, then you buy just one ticket to win; if there is little competition you may maximize your chances by purchasing many tickets; but if the competition is very high, it pays to save your money and buy tickets in a future raffle with improved odds.

Pizzari *et al.* also showed that cockerels were sensitive to the female's comb size, which the authors found to be a predictor of the female's average egg size and hence reproductive 'value'. Thus, males invested more of their sperm budget in females that had larger combs. Similarly, males were sensitive to female novelty. Over successive matings with the same female, a male's ejaculate size progressively fell — indicative of declining sperm reserves. But if a new female was presented to the same male, mating resumed with renewed enthusiasm, and ejaculate size returned to previous high sperm numbers. This reproductive response

Microbiology

Destructive approach

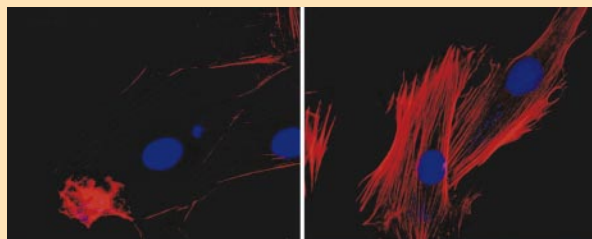
In the latest example of the wiliness of infectious bacteria, Tomoko Kubori and Jorge E. Galán have shown that *Salmonella typhimurium* exploits its host's protein-degradation machinery in order to reverse the substantial cellular reorganization that allows the bacterium to gain entry (*Cell* **115**, 333–342; 2003).

A common cause of food poisoning in humans, *S. typhimurium* eases its passage into intestinal cells by injecting various proteins into them, including one called SopE. This protein causes rearrangements in the cell's actin filaments — part of the intracellular 'skeleton' — and ruffling of the cell membrane, events that enable the bacteria to gain entry. The left image here shows the rearranged actin filaments (red), along with internalized bacteria (blue). Shortly afterwards, another injected

bacterial protein, SptP, counters these effects, restoring the status quo (right-hand image).

So how does *S. typhimurium* ensure that these events happen in the right order? One possibility is that it injects SopE first and SptP later. But Kubori and Galán rule out that idea: they show that SptP is present at the same levels as SopE within about 15 minutes of infection — just after the first evidence of actin rearrangement is seen. This also shows that SopE can produce these rearrangements even when SptP is about.

The authors find that levels of SopE subside later in the infection, when the actin reorganization ceases. But SptP is still easily detected. What causes the difference? Again, the authors eliminate the possibility that the proteins are produced to different extents in the infected cell. So they



looked at whether differential degradation might be the key. They find that it is. The main cellular machinery for chewing up unwanted or damaged proteins is the proteasome; when a proteasome inhibitor is added to infected cells, the levels of SopE remain high.

Why is SopE destroyed so quickly after infection whereas SptP is not? Kubori and Galán find that this is not due to the presence of other bacterial factors, so it must be determined by something intrinsic to the proteins

themselves. Previous work led the authors to suspect that a particular region of each protein, known as the secretion and translocation domain, controls their degradation. To test this, they swapped these portions of the proteins. The result was that SopE was treated like SptP, and vice versa — the actin rearrangements were not reversed in either case, because SptP was degraded quickly, and SopE was not. So what's the difference between these regions of each protein? That's a question for the future.

Amanda Tromans

was first described as the 'Coolidge effect'¹³ after US President Calvin Coolidge, who, while visiting a government farm, professed amazement at the prolific frequency of one cockerel's mating — until he was informed that each mating took place with a new female. The Coolidge effect has previously been documented only as a re-initiation of sexual behaviour¹⁴. Pizzari *et al.* show that new females resuscitate both mating behaviour and ejaculate size.

Pizzari and colleagues' results⁴ illustrate the selective forces that could be driving the evolution of male strategies for winning sperm competitions. However, it will be important to quantify the actual fertilization and reproductive rates of competing males — to see whether the 'spend-and-save' strategy is truly adaptive, translating into long-term or lifetime reproductive success. We also need more detailed information on the energetic and reproductive costs of producing sperm. Another question, given that the probability of encountering future receptive females is implicit in prudent sperm-investment strategies, is whether (as would be predicted) males show sensitivity to local cues of the abundance of receptive females when mating.

Finally, future research might also explore the mechanisms behind the modulation of ejaculate size. Physiological work on the promiscuous rodent *Peromyscus maniculatus* shows that muscular contractility of the vasa deferentia is directly sensitive to opioid hormones¹⁵. The vasa deferentia transport

sperm to the site of ejaculation, and opioid hormones are secreted during male–male interactions as direct cues of sperm competition. It seems logical that a similar kind of mechanism might operate in fowl, too. ■

Matthew J. G. Gage is in the School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK.
e-mail: m.gage@uea.ac.uk

1. Parker, G. A. *Biol. Rev.* **45**, 525–567 (1970).
2. Birkhead, T. R. & Møller, A. P. *Sperm Competition and Sexual Selection* (Academic, London, 1998).
3. Gage, M. J. G. & Morrow, E. H. *Curr. Biol.* **13**, 754–757 (2003).
4. Pizzari, T., Cornwallis, C. K., Løvlie, H., Jakobsson, S. & Birkhead, T. R. *Nature* **426**, 70–74 (2003).
5. Setchell, B. P. in *Reproduction in Mammals* Vol. 1, *Germ Cells*

and Fertilization (eds Austin, C. R. & Short, R. V.) 63–101 (Cambridge Univ. Press, 1982).

6. Olsson, M., Madsen, T. & Shine, R. *Proc. R. Soc. Lond. B* **264**, 455–459 (1997).
7. Parker, G. A., Ball, M. A., Stockley, P. & Gage, M. J. G. *Proc. R. Soc. Lond. B* **264**, 1793–1802 (1997).
8. Parker, G. A., Ball, M. A., Stockley, P. & Gage, M. J. G. *Proc. R. Soc. Lond. B* **263**, 1291–1297 (1996).
9. Gage, M. J. G. *Anim. Behav.* **42**, 1036–1037 (1991).
10. Wedell, N., Gage, M. J. G. & Parker, G. A. *Trends Ecol. Evol.* **17**, 313–320 (2002).
11. Martin, O. Y. & Hosken, D. J. *Anim. Behav.* **63**, 541–546 (2002).
12. Pilastro, A., Scaggiante, M. & Rasotto, M. B. *Proc. Natl Acad. Sci. USA* **99**, 9913–9915 (2002).
13. Wilson, J. R., Kuehn, R. E. & Beach, F. A. J. *Comp. Physiol. Psychol.* **56**, 636–644 (1963).
14. Dewsbury, D. A. *Psychol. Bull.* **89**, 464–482 (1981).
15. Pound, N. *Proc. R. Soc. Lond. B* **266**, 1755–1758 (1999).

Femtophysics

Birth of a quasiparticle

Alfred Leitenstorfer

As electronic devices shrink, the interaction between electrons and the silicon crystal lattice, described in terms of 'quasiparticles', is a central issue. Ultrashort laser pulses can track the birth of such a quasiparticle.

With few exceptions, electronic devices are based on silicon: using this semiconductor material, an extremely high density of basic switching elements (transistors) can be incorporated in complex circuits. Soon the size of transistors on commercial silicon chips might drop below 10 nanometres¹. At this level of

integration, the physical dimensions of each element are similar to the distance between atoms in the material, and the conventional description of charge flow in electronic devices breaks down. On page 51 of this issue, Hase *et al.*² present an experiment that gives a flavour of the colourful dynamics that might be expected under such conditions.

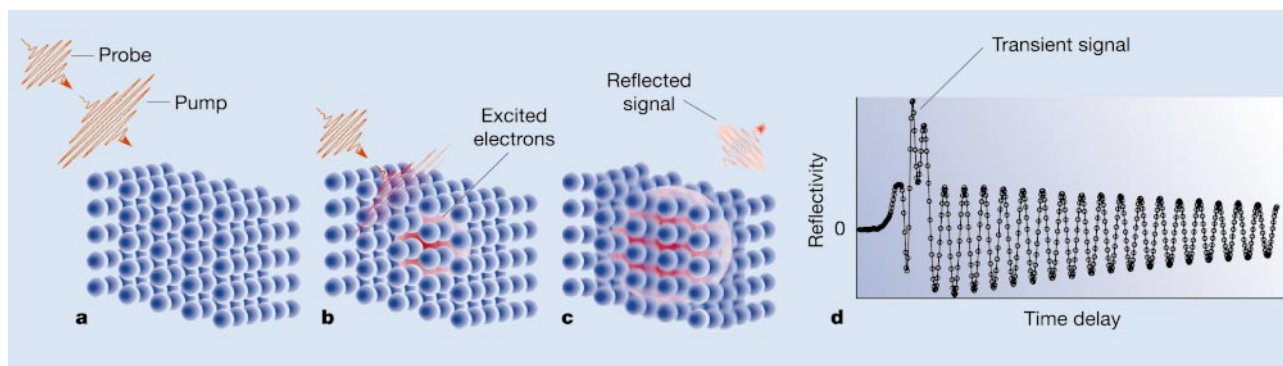


Figure 1 Caught in the act. a, Before excitation by a short, intense laser pulse (the pump), a pure silicon crystal contains few free electrons, and the atoms in the crystal lattice (blue) are virtually at rest. b, The pump hits the sample and creates a localized cloud of excited electrons, which interacts with the surrounding crystal lattice and causes characteristic vibrations. This process is described in terms of the generation of quasiparticles and is analogous to the abrupt injection of electrons into a nanometre-size

transistor. c, A delayed test pulse (the probe) is reflected from the sample surface and modified according to the instantaneous electronic conditions at the surface. d, By varying the time delay between pump and probe pulses, Hase *et al.*² have mapped out the excitation dynamics at the femtosecond timescale. The measured surface reflectivity shows a damped oscillation, and a transient signal lasting around 50 femtoseconds.

In a material such as silicon, the motion of electrons is influenced by the surrounding ion cores of the crystal lattice. These processes are complicated; but, on time and length scales that are much larger than molecular ones, they may be described in a straightforward way by assuming that the electron mass is altered and that its specific state of motion has a finite lifetime. In the language of condensed-matter physics, the electrons are 'renormalized' by the surrounding matrix, becoming 'dressed' as 'quasiparticles'. Until now, microelectronics engineers could quietly assume that interactions between these quasiparticles and their matrix occur instantaneously — that when an electron bumps against the ion lattice, there is a sudden change in its kinetic energy and momentum. But at nanometre scales, where this semiclassical picture is no longer applicable and quantum effects become relevant, it is not clear at present how to model the electron behaviour.

There is another complication too. In principle, it should take a finite amount of time to establish the quasiparticle in a non-equilibrium situation or to complete a scattering with the lattice. The typical timescale for this is given by the oscillation cycle of the lattice vibrations. These are quantized, and are given the name 'phonons'. As transistors shrink to nanoscale dimensions, they might respond so quickly to a pulse of electric current that their internal switching times approach this limiting timescale.

The build-up of quasiparticles has been observed³ in compound semiconductors such as GaAs. But quantum kinetic processes have not previously been accessible in silicon. To measure them requires extremely high temporal resolution, which can currently only be achieved using ultrashort laser pulses, of a few femtoseconds duration ($1\text{ fs} = 10^{-15}\text{ s}$). However, the electrons moving inside silicon are typically of low energy

and their coupling to the photons of the laser pulse is relatively complicated, so direct insight has been hampered.

Hase *et al.*² have worked around this problem by using femtosecond ultraviolet pulses first to excite the electrons to much higher energies, and then to probe the subsequent dynamics (Fig. 1). The absorption of high-energy photons by the electrons is a relatively simple process, and interpretation of these data is straightforward. The authors then trace what happens to the electrons using the electro-optic effect, which describes how the excited electrons and lattice vibrations change the optical properties of the material. These measurements are truly a *tour de force*: the electro-optic effect is absent in bulk silicon (because it is a centrosymmetric system); but Hase *et al.* have sampled reflected light from the silicon surface and show that it is modulated. Their apparatus needed to be a hundred times more sensitive than in previous experiments⁴ (on polar materials), because only a few atomic layers at the surface contribute to the signal.

The time traces recorded by Hase *et al.* consist of two parts (Fig. 1). One is an exponentially damped oscillation, which originates from the interactions of lattice phonons and the photo-generated electrons. The other is an aperiodic, transient signal from the electrons themselves that appears only for 50 fs. Other researchers might have been content to focus on the phonon oscillations alone. But with an experiment capable of 10-fs time resolution, Hase *et al.* realized that the electronic response does not simply follow the excitation pulse. Instead, it contains more information on the build-up, or dressing, of the electronic quasiparticles. Cleverly, they transformed their time-dependent data into time–frequency space, creating a sequence of spectra defined by varying the delay between the application of

the pump and the probe laser pulse. And there, they have found new physics.

Extremely close in time to the initial excitation, Hase *et al.* have uncovered signatures of the force exerted by the electronic charge on the lattice, and by the lattice on the developing quasiparticles. The most remarkable feature is a dip in their spectra where the electronic and lattice responses overlap in the time–frequency plane. The authors attribute this to a so-called Fano interaction⁵, originating from a coherent superposition of phonons and the broad continuum of electronic excitations. Solid-state physicists talk about 'quantum correlations' in this context; researchers from the field of quantum optics might use the term 'entanglement'. In any case, Hase *et al.* have watched, step by step, the ultrafast birth of quasiparticles in silicon, as electrons are dressed by the lattice.

This paper² stimulates many fascinating questions, some general in nature, others specific to future research. For example, specialists might wonder how theory will succeed in explaining quantitatively the complicated phenomena that are seen in this experiment. Also, how closely does the behaviour of the very energetic electrons observed by Hase *et al.* mimic that of their lower-energy counterparts travelling through silicon chips? The day will come when quantum physics directly influences the functionality of computers and other electronic equipment that we use in everyday life — the question is, when? ■

Alfred Leitenstorfer is in the Department of Physics, University of Konstanz, PO Box M 695, D-78457 Konstanz, Germany.

e-mail: alfred.leitenstorfer@uni-konstanz.de

- Doyle, B. *et al.* *Intel Tech. J.* **6**, 42–54 (2002); www.intel.com/research/silicon
- Hase, M., Kitajima, M., Constantinescu, A. M. & Petek, H. *Nature* **426**, 51–54 (2003).
- Huber, R. *et al.* *Nature* **414**, 286–289 (2001).
- Cho, G. C., Kütt, W., Kurz, H. & Ibach, H. *Phys. Rev. Lett.* **65**, 764–766 (1990).
- Fano, U. *Phys. Rev.* **124**, 1866–1878 (1961).

Neurobiology

Signals that make waves

Louis F. Reichardt

Neurons in the brain release proteins called neurotrophins, which bind to glial cells and unleash a wave of calcium ions inside them. This could be the missing link in a communication circuit between glia and neurons.

Our brains contain billions of nerve cells, woven together in an intricate network. But another class of brain cells, glial cells, vastly outnumbers the neurons. These cells provide essential support, forming a scaffold to hold the neurons in place, insulating the neuronal protrusions, providing nutrients and clearing debris. On page 74 of this issue, Rose *et al.*¹ reveal another link between glial cells and neurons. Neurotrophins — proteins that bind to and promote the survival of neurons — also bind to astroglial cells. Rose *et al.* find that this binding triggers the release of calcium ions inside these cells. This might regulate communication between neurons.

Neurotrophins are crucial for the normal development and functioning of the brain. These proteins bind directly to neurons, triggering intracellular signals that culminate in survival, growth or the modulation of neuronal function. About a decade ago, the major receptors to which neurotrophins bind on neurons were shown to be a sub-family of 'receptor tyrosine kinases', named Trks (ref. 2). These receptors span the neuronal membrane; neurotrophin binding at the outside of the cell causes them to dimerize, which in turn activates the intracellular tyrosine-kinase portion of the receptor. This portion then phosphorylates other intracellular proteins, initiating signalling cascades

inside the neurons. These lead to changes in protein function and gene expression that promote neuronal survival and maturation. The Trk receptors can also directly regulate other proteins in the neuronal membrane, including ion channels and receptors for neurotransmitters, which are essential for neuronal function^{3,4} (Fig. 1a).

There are three *trk* genes, and each encodes more than one protein product. For instance, the *trkB* and *trkC* genes encode not only full-length TrkB and TrkC proteins, but also truncated forms lacking the intracellular tyrosine-kinase domain. However, little was known about the function of the truncated forms, as most studies of Trk receptors had involved the full-length, kinase-containing forms, which are abundant in neurons. The truncated forms are found mainly in glial cells, which lack the kinase-containing receptors. And the truncated form of TrkB is widely expressed both during development and in the adult brain.

What is the function of these truncated receptors, and why are they present on glial cells? Some clues, but no clear answers, have come from several studies. As the brain matures, the ratio of truncated to kinase-containing forms of TrkB increases dramatically in the outer layer of the brain (the cerebral cortex), indicating that the functions of the truncated receptors are developmentally

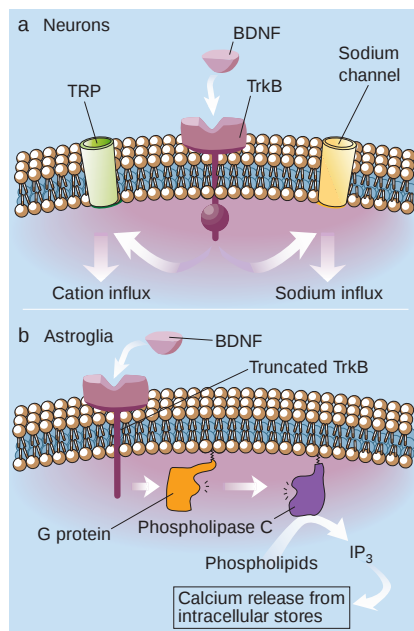


Figure 1 A multifunctional receptor. The TrkB receptor has various ways of regulating the ion concentrations in different brain cells. **a**, The full-length receptor occurs abundantly on the surface of neurons. When it binds its ligand, BDNF, the receptor's intracellular 'kinase' domain phosphorylates the enzyme phospholipase C, activating the transient receptor channel TRP. TrkB can also activate a sodium channel (Na(v)1.9) via a faster but less well understood interaction that probably does not require kinase activity. **b**, Astroglia contain a truncated form of the TrkB receptor (TrkB-T1). Its function was unclear, but Rose *et al.*¹ now show that BDNF binding to this receptor activates the release of calcium from intracellular stores, through a signalling pathway that involves an as yet unidentified 'G protein'. This in turn promotes the generation of inositol trisphosphate (IP₃) through phospholipase C. When IP₃ binds to its intracellular receptor, calcium is released.



100 YEARS AGO

It is announced by the *Electrician* that this year it is proposed to award the Nobel physics prize to Signor Marconi, the chemistry prize to Prof. Arrhenius, and the medicine prize to Prof. Finsen. Each prize is worth about 8000l. From *Nature* 5 November 1903.

50 YEARS AGO

The Nobel Prize for Medicine and Physiology for 1953 has been awarded jointly to Prof. H. A. Krebs, professor of physiology in the University of Sheffield and director of the Medical Research Council Unit for Research in Cell Metabolism, and Dr. F. Lipmann, head of the Biochemical Research Laboratories, Massachusetts General Hospital. **Prof. H. A. Krebs, F.R.S.**: Prof. Krebs has mostly been concerned with the study of metabolic processes by experiments *in vitro*. The first of his two greatest contributions to biochemistry was from Freiburg in 1932, when he elucidated the mechanism of urea synthesis in the liver, by discovering the participation of ornithine, citrulline and arginine through a cyclical process — a concept of unprecedented nature. His subsequent observations on the deamination of amino-acids demonstrated D-amino-acid oxidase, and laid foundations for future studies of the L-acids. In Cambridge in 1935 he proved the synthesis of glutamine from glutamic acid and ammonia in tissue slices. After moving to Sheffield, he announced in 1937 his second major contribution, the citric-acid cycle.

Dr. Fritz Lipmann: After studying problems of muscle metabolism in Meyerhof's laboratory and of fermentation in the Carlsberg laboratory, Lipmann set the pattern for his future work when in 1937 he began to analyse the oxidation of pyruvate to acetate by bacteria. He found that the oxidation is accompanied by phosphorylation, and announced from Cornell University in 1939 that the 'energy-rich' ester acetyl phosphate is an intermediate... Moving to Boston, Lipmann realized that acetyl phosphate is not formed in pyruvate oxidation in animal tissues; some other substance had to be sought... Studying the biological acetylation of sulphanilamide, he discovered a new coenzyme, coenzyme A... Finding that it is a derivative of the vitamin pantothenic acid and a general constituent of living organisms, he quickly realized that it is of fundamental importance in carbohydrate and fat metabolism. From *Nature* 7 November 1953.

regulated⁵. Truncated TrkB and TrkC have also been shown to form dimers with, and thereby inhibit the activation of, the full-length forms of these receptors. And truncated TrkB can rapidly bind and internalize the neurotrophins BDNF and NT4, preventing their diffusion⁶. Studies of cultured astroglia (matrix-forming glial cells in the brain) and Schwann cells (glial cells that insulate nerve-cell projections in the peripheral nervous system) have also revealed that truncated TrkB promotes both the internalization and subsequent release of BDNF. This suggests that these cells could act as a reservoir of BDNF to promote the growth and maturation of adjacent neurons⁷. Finally, several groups have shown that the truncated receptors can, through more direct mechanisms, regulate the neurons' intracellular pH and maturation, but what these mechanisms might be has been somewhat enigmatic^{8–10}.

Rose *et al.*¹ now take a step towards solving this enigma. They show that when BDNF is applied to cultures of astroglia and slices of brain, it generates waves of calcium release inside the glial cells. Calcium is an important 'second messenger' that regulates the release of neurotransmitters and other molecules from cells, as well as controlling the activities of several signalling proteins and ion channels. BDNF is released by neurons and is present in the brain at the sub-nanomolar concentrations required to generate calcium waves, so this protein might be a means by which neurons signal to glia.

The authors show that the BDNF-induced production of calcium waves in astroglia requires truncated ('T1'), but not kinase-containing, TrkB. And wave generation requires intracellular but not extracellular calcium — waves are prevented by inhibiting a receptor, the inositol trisphosphate receptor, through which calcium is released from intracellular compartments. Rose and colleagues' study also implicates several other signalling molecules in the pathway from receptor activation to wave generation (Fig. 1b). These include an as yet unidentified G protein (a class of protein that often couples cellular receptors to intracellular signalling pathways) and phospholipase C (an enzyme that generates inositol trisphosphate from phospholipids). It will be important to identify the G protein involved in this signalling cascade.

What is the significance of the waves of calcium release in astroglia? In so-called pyramidal neurons of the cerebral cortex, both the expression and release of BDNF are induced by calcium and electrical activity. And in adjacent astroglia, increased levels of calcium evoke the release of the excitatory neurotransmitter glutamate, enhancing the activity of the synapses formed by neighbouring neurons¹¹. (Synapses are the connections through which electrical impulses can pass from one neuron to the next.) Calcium waves can also propagate through many

astroglia and regulate synapses over wide areas by means of gap junctions, which are connections between cells that allow molecules to move from one cell to another¹¹.

It was known that BDNF promotes both short- and long-term enhancement of synaptic strength, but the underlying mechanisms were poorly understood. Rose and colleagues' findings¹ suggest that it may exert its effects on neuronal synapses in part by triggering calcium signalling in astroglia. The physiological functions of this signalling pathway, and which of the other truncated forms of TrkB and TrkC can initiate it, remain to be discovered. ■

Louis F. Reichardt is in the Neuroscience Program, Department of Physiology, and Howard Hughes

Medical Institute, University of California at San Francisco, San Francisco, California 94143-0723, USA.

e-mail: lfr@cgl.ucsf.edu

1. Rose, C. R. *et al.* *Nature* 426, 74–78 (2003).
2. Huang, E. J. & Reichardt, L. F. *Annu. Rev. Neurosci.* 24, 677–736 (2001).
3. Chao, M. V. *Nature Rev. Neurosci.* 4, 299–309 (2003).
4. Blum, R., Kafitz, K. W. & Konnerth, A. *Nature* 419, 687–693 (2002).
5. Allendoerfer, K. L. *et al.* *J. Neurosci.* 14, 1795–1811 (1994).
6. Biffo, S., Offenhauser, N., Carter, B. D. & Barde, Y. A. *Development* 121, 2461–2470 (1995).
7. Alderson, R. F., Curtis, R., Alterman, A. L., Lindsay, R. M. & DiStefano, P. S. *Brain Res.* 871, 210–222 (2000).
8. Hapner, S. J., Boeshore, K. L., Large, T. H. & Lefcort, F. *Dev. Biol.* 201, 90–100 (1998).
9. Yacoubian, T. A. & Lo, D. C. *Nature Neurosci.* 3, 342–349 (2000).
10. Baxter, G. T. *et al.* *J. Neurosci.* 17, 2683–2690 (1997).
11. Haydon, P. G. *Nature Rev. Neurosci.* 2, 185–193 (2001).

Ecology

Palms in motion

Peter D. Moore

Being eaten alive then dumped with the resulting droppings can be all to the good — if you are a palm fruit in the Amazonian tropical forest, that is, and the consumer is a large mammal.

Getting away from one's parent has its advantages. When animals disperse plant seeds, for instance, the seeds avoid competing with their well-established parent and also escape the organisms that home in on the ready source of nutrition that an older tree represents. When the animal is large, and the seeds are ingested, the distances covered can be considerable. And the seeds conclude their journey securely buried in faeces that may further protect them from parasites and predators.

The dispersal of palms by tapirs in the forests of Brazil, described by José Fragoso and colleagues in *Ecology*¹, provides a good example of this process and also explains aspects of the patterns in which palms grow.

Most models of seed dispersal, based on direct demographic studies, tend to consider long-distance dispersal as relatively rare, and as insignificant². Genetically based research, in contrast, recognizes its importance in the spread of genes³. A model that has proved most useful for tree-seed dispersal in tropical



Figure 1 Passing through: the Brazilian tapir, *Tapirus terrestris*.

MICHAEL & PATRICIA FOGDEN/CORBIS

Cell biology

Enlightened messages

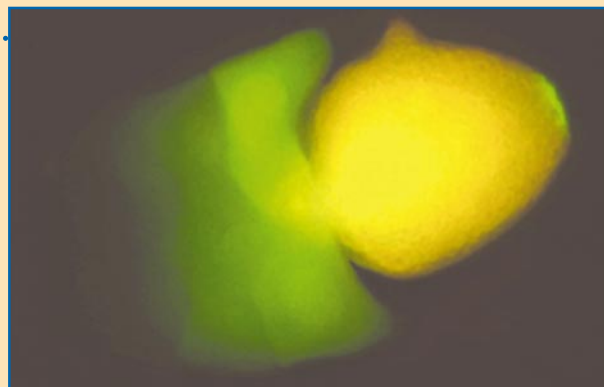
Writing in *Proceedings of the National Academy of Sciences*, Diana Bratu and colleagues describe a way to trace the movement of messenger RNA (mRNA) in living cells (doi:10.1073/pnas.2233244100).

Every moment inside cells, thousands of proteins and RNA molecules shuttle from one location to another. By fusing cellular proteins to other, fluorescent proteins, their movements can be followed visually. But it is not possible to tag RNA in the same way. Bratu *et al.* have devised a different sort of tracking system to monitor the transport of specific RNA messages. They created fluorescent 'molecular beacons' — short stretches of nucleic acids

that seek out and bind to complementary mRNA sequences.

Attached to one end of the beacon is a fluorophore; a fluorescence quencher is fixed to the other end. The single-stranded beacon normally folds back on itself, forming a double-stranded hairpin structure in which the quencher and fluorophore are held in close proximity. But when the beacon binds to its complementary sequence in the RNA message, it unfolds — the fluorophore is separated from the quencher and the mRNA lights up.

To test their tracking system, the authors designed beacons for the *oskar* mRNA of fruitflies. The *oskar* message encodes the Oskar protein,



which is involved in patterning the developing fruitfly egg. It is produced in 'nurse cells', which nurture the developing eggs. After entering and traversing the egg, *oskar* mRNA accumulates at the posterior end.

Bratu *et al.* injected *oskar*-specific molecular beacons into a living egg to see whether the pattern

of fluorescence would unfold as expected. As the picture shows, it did — the authors detected the *oskar*-specific signal (green) in the neighbouring nurse cells and at the posterior end of the egg, whereas the signal from a control beacon was located throughout the egg (yellow).

Clare Thomas

forest is that of Janzen⁴. In this, recruitment of individuals to a community is regarded as a consequence of dispersal from the parent plant, combined with predation by seed-eating organisms such as bruchid beetles, which are particularly adept at dealing with the many secondary plant compounds that deter most seed predators.

Because seed predators tend to concentrate around the parent tree, recruitment there is low, despite a high input of new seeds. There is a critical distance from the parent — within the range of frequent seed arrival, but with few predators — in which recruitment is most abundant. Beyond this, seed arrival is limited by dispersal ability. This model accounts for why many species of tree tend to be widely separated in tropical forests. But it does not explain why others may occur in clumps.

Working on Maracá Island in the Amazon basin of Brazil, Fragoso *et al.*¹ studied the effects of tapir feeding on the fruits of the palm *Attalea maripa* (formerly known as *Maximiliana maripa*), which is unusual in that it tends to grow in clusters. The fruits are 5–8 cm in length and are produced in large numbers. They consist of a fibrous husk, a layer of yellow pulp and a further woody layer — the endocarp — that encloses the seeds. Many small mammals feed upon the fruits, first removing the husk and then eating the pulp. The inner part is then abandoned and is usually attacked by bruchid beetles, which lay their eggs on the damaged fruits: their larvae penetrate to the seeds and consume them. Infestation close to the parent tree can approach 100%.

Tapir fruit-eating behaviour is different from that of the smaller mammals. These are

large animals (about 250 kg; Fig. 1), with home ranges of several thousand hectares. They ingest the palm fruits intact, digest the pulp, and finally pass the endocarps and husk fragments in their faeces. They habitually defecate in specific latrine sites, both in the upland, drier areas and in the wetland swamps of Maracá Island. The endocarps at these sites are 98% viable, and the establishment of young palms is far more successful there than around the parent trees. Aerial surveys confirmed the existence of aggregations of palms, possibly reflecting former patterns of tapir latrines. But to find out more about the factors enhancing seed survival, experiments were required.

Fragoso *et al.* tackled matters as follows. They collected palm endocarps, placing them at different distances from the parent trees, and enclosing them in wire frames to exclude all large predators but allow access by beetles. Some plots contained clean endocarps. In others, the endocarps were covered with tapir faeces (from which any ingested endocarps had been removed) to simulate the burial conditions at latrine sites.

The results showed that, up to a point, survival was significantly enhanced by distance from the parent site (a consequence of the scarcity of bruchid beetles away from those sites). Survival was also greater when endocarps were buried in faecal material (making it more difficult for the beetles to locate the endocarps and lay their eggs on them). The effect of burial, however, became insignificant in very distant sites, presumably also because of the scarcity of the seed predator. Modelling of palm population dynamics under these conditions will evidently prove complex⁵.

This work¹ illustrates the complexity of interactions between palm, tapir and beetle. But it also has a bearing on conservation. Restriction of tapir movement, by habitat fragmentation for example, could severely affect palm population dynamics. The patches of palm in the forest and sometimes also in the savanna, created at tapir latrines, constitute a mosaic of habitats in which biodiversity thrives. Such interactions between plants and animals might be a central determinant of the rates of adjustment of vegetation to changing conditions — the future of the forest could in part lie within the intestines of a tapir.

Peter D. Moore is in the Division of Life Sciences, King's College, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NN, UK. e-mail: peter.moore@kcl.ac.uk

1. Fragoso, J. M., Silvius, K. M. & Correa, J. A. *Ecology* **84**, 1998–2006 (2003).
2. Cain, M. L., Milligan, B. G. & Strand, L. E. *Am. J. Bot.* **87**, 1217–1227 (2000).
3. Rousset, F. in *Dispersal* (eds Clobert, J., Danchin, E., Dhondt, A. A. & Nichols, J. D.) 18–28 (Oxford Univ. Press, 2001).
4. Janzen, D. H. *Am. Nat.* **104**, 501–528 (1970).
5. Turchin, P. *Complex Population Dynamics: A Theoretical/Empirical Synthesis* (Princeton Univ. Press, 2003).

Editorial note

The News and Views article "Molecular biology: Complicity of gene and pseudogene" (*Nature* **423**, 26–28; 2003) discussed the discovery of pseudogene function in the mouse, as described by S. Hirotsume *et al.* on pages 91–96 of the same issue. There is an earlier report of pseudogene function, in a mollusc and with a different mechanism (S. A. Korneev, J.-H. Park & M. O'Shea *J. Neurosci.* **19**, 7711–7720; 1999).

Medicine

Cancer-killing virus on a drip

Cancer Cell **4**, 263–275 (2003)

There is a new virus on the block that kills tumours but not their host, according to David F. Stojdl and colleagues.

Budding tumours switch off from many signals around them; for example, they often become indifferent to interferon, a major regulator of cell growth and death. But this so-called cytokine also coordinates a cell's antiviral response, making tumour cells with disabled interferon signalling vulnerable to a viral attack. It is thus feasible — in theory — to eliminate tumours by unleashing a virus. But it's no good if the patient then succumbs to a raging viral infection.

Stojdl *et al.* have now identified variant strains of the vesicular stomatitis virus (VSV) that kill the tumour and spare the host. Following simple intravenous delivery in mice, the virus destroyed tumours — even metastases — and was rapidly cleared by the mice thereafter. What's the secret?

To evade an antiviral attack by immune cells, ordinary VSV also tampers with interferon signalling in the cell. Stojdl *et al.* show that their mutant strains lack the capacity to do so. Thus, they induce a powerful 'cytokine cloud' wherever they go, and a proper antiviral response in healthy cells. Tumour cells are oblivious to this cloud and are eliminated one by one by the virus.

Marie-Thérèse Heemels

Physiology

Hotter equals smaller

Proc. R. Soc. Lond. B doi:10.1098/rspb.2003.2538 (2003)

Cold-blooded animals such as fish show an inverse relationship between the temperature at which they are reared and their final body size — something known as the temperature–size rule. Now David Atkinson *et al.* show that single-celled organisms called protists follow a similar rule.

The researchers carried out a meta-analysis of published data on water-dwelling protists, including amoebae, ciliates, diatoms, dinoflagellates and flagellates. By combining 65 data sets, they found that the organisms' average size decreases by 2.5% for every 1 °C boost in temperature above 15 °C.

The team proposes two hypotheses to explain the size trend. The availability of respiratory gases such as oxygen and carbon dioxide could limit growth as the temperature rises, because the amount of gases dissolved in water decreases, whereas the organisms' demand for such gases rises because of their increased metabolic rate. Alternatively, the organisms may gain a competitive edge by dividing earlier as population growth increases.

The equation can be used to better estimate how the number or biomass of protists changes with temperature, and this might be useful in modelling aquatic ecosystems.

Helen Pearson

Glaciology

Thin ice

Science **302**, 856–859 (2003)

In two abrupt episodes during the past ten years, the Larsen Ice Shelf on the Antarctic Peninsula lost its northern-most sections — Larsen A and Larsen B — when they fragmented into icebergs. What caused the break-up, and what does it imply for the much larger Larsen C shelf that makes up the rest of this sea-ice sheet? Andrew Shepherd and colleagues say that the Larsen Ice Shelf has been thinning steadily since at



least 1992, and suspect that this contributed to the weakening of Larsen A and B, which together had an area of more than 5,000 km².

The authors studied nine years' worth of satellite altimeter measurements of the ice sheet's surface height, and find that, since 1992, Larsen B and C have decreased in height at annual rates of about 0.17 and 0.08 metres, respectively. These decreases do not necessarily imply melting-induced mass loss: some of the change might have come from increased densification of the ice. Yet that alone is unlikely to account for all of the thinning; the researchers say that there has also been melting of basal ice. If their estimate of the ice erosion rate is correct, within 100 years Larsen C will be close to the thickness at which Larsen B broke up.

Philip Ball

Prion diseases

Untangling toxicity

Science **302**, 871–874 (2003)

Giovanna Mallucci *et al.* have found a way to reverse the first signs of brain destruction in a mouse model of prion disease.

The disease-associated prion protein

causes normal versions of the protein to change shape and stick together in the brain. As the prion clumps accumulate, neurons die, leaving the brain riddled with holes like a sponge. But what kills the neurons? To investigate this, Mallucci *et al.* created genetically engineered mice that produce the normal prion protein in neurons during the first 10–11 weeks of life only. Older mice continue to produce this protein in other brain cells.

The authors injected abnormal prions into the brains of the mice at 1 week old. By 8 weeks, all the mice had developed early signs of 'spongiform' disease. Remarkably, these signs disappeared in 12-week-old mice, when neurons were no longer making the normal protein. And the animals remained disease-free for at least 48 weeks, even though prion clumps were still being produced by other brain cells.

Mallucci *et al.* suggest that conversion of the normal protein to the disease-related form is toxic only if it occurs within neurons. It remains to be seen whether eliminating normal prion proteins in human neurons could be equally beneficial.

Clare Thomas

Cell biology

Arrested replication

J. Cell Biol. **163**, 245–255 (2003)

Cells use several mechanisms to ensure that, when DNA is damaged, it is repaired before it can be replicated and produce mutations in daughter cells. Matthew P. Stokes and W. Matthew Michael now describe a previously unsuspected pathway that prevents global DNA replication when damage is present.

DNA damage is known to activate factors called checkpoint proteins, which prevent new replication from being initiated. But it was unclear whether damaged DNA itself is sensed by checkpoint proteins, or whether ongoing replication stalls at the site of damage, forming abnormal 'replication intermediates' that are recognized by the checkpoint machinery. Using frog extracts, Stokes and Michael find that the presence of damaged plasmids (circular DNA molecules) causes replication to stall on separate, undamaged chromosomes. This suggests that the damaged DNA activates a diffusible replication inhibitor that can act on other DNA molecules. The inhibitor prevents DNA from binding to PCNA, a complex that tethers DNA-replicating enzymes to chromosomes. Unexpectedly, activation of the inhibitor is independent of the DNA-damage checkpoint, but causes checkpoint activation on undamaged DNA.

The nature of the inhibitor is unknown, but preliminary data suggest that it does not bind to PCNA directly. Instead, it may bind 'factor X' — a hypothetical protein that loads PCNA onto DNA.

Angela K. Eggleston

Crop freezes and land-use change in Florida

Draining the state's southern wetlands may have raised the incidence of harmful frosts.

South Florida experienced a significant change in land usage during the twentieth century, including the conversion of natural wetlands into agricultural land for the cultivation of winter vegetable, sugar cane and citrus crops. This movement of agriculture from more northerly areas was intended partly to escape the risk of damaging winter freezes. Here we present evidence from a case study using a coupled atmosphere and land-surface computer-modelling system that suggests that the draining of wetlands may have inadvertently increased the frequency and severity of agriculturally damaging freezes in the south of Florida.

On 19 January 1997, a rare freeze inflicted severe damage in agricultural areas of south Florida that were once natural wetlands, with below-freezing temperatures extending to the tip of the peninsula. This event, chosen here for our case study, resulted in losses in the fresh-vegetable and sugar-cane sectors that alone exceeded US\$300 million¹. Furthermore, nearly 100,000 migrant farm workers were displaced or unemployed as a result of the freeze².

We used the Regional Atmospheric Modelling System (RAMS)³, a comprehensive meteorological modelling system that includes a sophisticated land-surface scheme to represent the effects of surface properties on atmospheric processes⁴, to investigate the impact of anthropogenic changes in land coverage on this freeze. A pair of simulations was undertaken in which the model configuration was identical, except that in one simulation the data represented pre-1900s (almost natural) land cover, whereas in the other they represented 1993 (near-present-day) land usage. These data sets (Fig. 1a, b) reflect the conversion of natural wetlands to agricultural land in the areas of south Florida that were affected by the freeze.

In key agricultural areas that were once natural wetlands, particularly the areas used for high-density cultivation of winter vegetables, sugar cane and citrus fruits to the south and southwest of Lake Okeechobee, the simulation incorporating current land coverage produced minimum temperatures that were generally colder (Fig. 1c) and were below freezing for a longer period (Fig. 1d) than that using natural land coverage. The results reveal that when land-surface properties were specified to represent natural land cover, a persistent heat flux from wetlands was sufficient to hold the simulated temperature above freezing throughout the night in many of these areas (results not shown).

In other agricultural areas of south Florida that were once natural wetlands, such as portions of the Kissimmee Valley (north of Lake Okeechobee), the model simulated a freeze regardless of the specification of land-surface properties. However, even in these areas, minimum temperatures were generally colder and were below 0°C for a longer period when present-day land use was prescribed (Fig. 1c, d). The duration of exposure to sub-zero temperatures is critical in determining the amount of crop damage that occurs during a freeze⁵.

Our results indicate that, even in areas where a freeze would have occurred irrespective of land-surface properties, the agricultural damage may have been worse than it would have been if natural wetlands had still been present in those areas. Results were similar when the same modelling method was applied to other recent agriculturally damaging freezes in south Florida, including

the events of 26 December 1983 and 25 December 1989.

We have shown that the likelihood of agriculturally damaging freezes in south Florida has increased as a result of the conversion of its natural wetlands to agriculture. This is ironic, considering that the move was instigated partly to avoid the freezes that occur further to the north^{6–8}. Our results provide another example of the potential for anthropogenic changes in land usage to perturb the climate system⁹.

Curtis H. Marshall*, **Roger A. Pielke Sr***, **Louis T. Steyaert†**

*Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado 80523, USA
e-mail: curtis@atmos.colostate.edu

†US Geological Survey, EROS Data Center and NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

1. Economic Research Service, US Department of Agriculture
Agricultural Outlook 2 (1997).

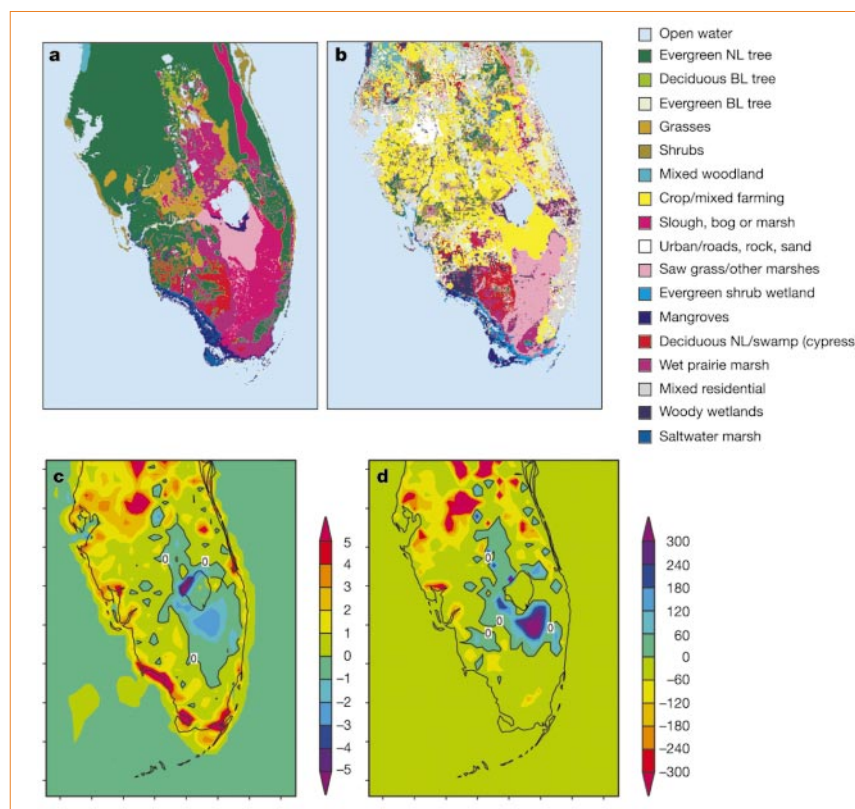


Figure 1 Minimum temperatures, and the duration of periods of sub-zero temperatures, in areas cultivated from drained wetlands in south Florida. **a, b**, Classes of land use specifying model properties for simulations incorporating land-surface conditions before the 1900s (**a**) and in 1993 (**b**). NL, needleleaf; BL, broadleaf. **c**, Difference between the model's simulated minimum temperatures (in °C) near ground level on 19 January 1997; differences were determined as the values derived from 1993 usage minus those from pre-1900s usage. Locations inside the zero contour experienced colder minimum temperatures when 1993 land use was used in the model. **d**, Difference in the duration (in min) of subzero-temperature periods for the two different model simulations; differences were determined as in **c**. Areas inside the zero contour experienced freezing temperatures for longer when 1993 land use was used in the model.

2. Rural Migration News Florida Freezes; West Floods 3 (1992).
3. Pielke, R. A. et al. Meteorol. Atmos. Phys. 49, 69–91 (1992).
4. Walke, R. L. J. Appl. Meteorol. 39, 931–944 (2000).
5. Wolford, L. V. Weekly Weather Crop Bull. 42, 7–8 (1955).
6. Whittaker, H. M. Proc. Florida State Hort. Soc. 98, 46–48 (1985).
7. Miller, K. A. Clim. Res. 1, 133–144 (1991).
8. Attaway, J. A. A History of Florida Citrus Freezes (Florida Science Source, Lake Alfred, 1997).
9. Kalnay, E. & Cai, M. Nature 423, 528–531 (2003).

Competing financial interests: declared none.

Arachnology

Scavenging by brown recluse spiders

The brown recluse spider (*Loxosceles reclusa*) is a threat to humans and establishes huge populations in urban habitats throughout central North America^{1,2} — more than 2,000 of these spiders were recorded in a single house in Kansas³. What do these spiders eat in order to build and maintain such numbers? Here I combine laboratory prey-choice experiments with observations of the behaviour of *L. reclusa* in houses to show that this spider prefers dead, scavenged prey over live prey. This finding may explain how immense populations of these spiders can flourish even in adverse conditions.

All spiders consume live prey and they have evolved a diverse array of adaptations to enhance its detection and capture. Some spiders also supplement their diets with unusual food sources such as pollen⁴, nectar⁵ and insect eggs⁶. However, spiders would not be expected to prefer dead prey, given the stimulus of catching live prey that results from their extreme sensitivity to even the slightest movement of their victim.

Bites from brown recluse spiders (Fig. 1) can constitute a medical emergency in humans, inflicting slow-healing necrotic wounds and extensive tissue damage^{1,2}. The same powerful paralyzing venom that causes these effects in humans also allows the spider to immobilize its prey and avoid being injured itself⁷. Unlike most wandering spiders, which use vision, stealth and strength to capture prey, *L. reclusa* attacks, retreats and then returns to feed at its convenience⁷; it has relatively poor vision⁷ and does not readily respond to substrate or airborne vibration.

I investigated the predatory behaviour of *L. reclusa* in 71 homes in Kansas and did not see spiders catching live prey — indeed, the spiders actually fled from potential prey. However, I witnessed the spiders locating and consuming dead prey, without prior attack, in more than 25 houses.

For the prey-choice experiments, I placed adult male and female *L. reclusa* ($n = 147$) into individual plastic boxes (12 × 17 × 6 cm) and fed them a variety of prey, both dead and alive, once a week, and then starved them

for two weeks. The spiders were subsequently presented with equally sized live and dead prey (dead prey was killed by freezing at -80°C). Spiders were kept in the dark and observed under low light every hour for evidence of prey choice, which was verified by fang penetration and feeding for 5 min or longer.

In these experiments, 81.4% of *L. reclusa* chose dead over live waxmoth larvae (*Achroia grisella*; $n = 59$), 75.6% chose dead over live domestic crickets (*Acheta domestica*; $n = 41$) and 97.6% chose dead over live yellow mealworm larvae (*Tenebrio molitor*; $n = 41$). Overall, spiders chose dead prey in 84.4% of trials ($\chi^2 = 72$, $P < 0.001$), indicating a clear preference for scavenging dead prey.

To test the effects on the spiders of eating dead rather than live prey, I used three groups of ten spiders that had been respectively given: old dead crickets that had been envenomated and partially digested by other *L. reclusa* two weeks previously; German cockroaches (*Blattella germanica*) that had been dead for at least a month; and *B. germanica* that had been killed 24 h earlier with Cessco-5E insecticide (0.5% pyrethrin). All spiders consumed the prey within 24 h of its being introduced, and showed their resilience in that no obvious negative effect was manifest over the ensuing 10 months, during which a regular feeding regime was followed.

The feeding preferences that I observed in urban habitats are consistent with my findings with captive spiders, indicating that *L. reclusa* actively searches for dead prey and ignores live prey. Spiders will even remain motionless and allow their prey to walk over them without attacking it, such is the extent to which they prefer dead prey. In five *T. molitor* trials, spiders attacked and killed live prey but did not eat it, yet consumed prey killed by freezing. Capture of live prey (15.6%) was often the result of prey walking into spiders rather than of active hunting.

The sequence of predatory behaviour shown by *L. reclusa* (attack, retreat and feed later) means that prey may escape before it is relocated. Spiders sometimes leave partially eaten carcasses that other spiders subsequently find and consume. It is likely that dead prey provides an easily accessible source of nutrition without incurring the additional costs or risks associated with attacking and manipulating live prey.

L. reclusa is an opportunistic feeder rather than an obligate predator or obligate scavenger, but it prefers dead over live prey. In an environment such as a house, opportunists and scavengers have an advantage over more selective predators because their feeding requirements are more easily met⁸. Insects are attracted to houses by light, food and shelter, as well as for other unknown reasons, but they can easily die there from starvation, desiccation, overexertion, pesticide



Figure 1 Still watches: the brown recluse spider *Loxosceles reclusa* prefers prey that is motionless or already dead, feigning indifference itself to avoid a struggle with living prey.

exposure or other causes. I conclude that scavenging contributes significantly to the diet of *L. reclusa* and may be important for the survival of this species in natural habitats, as well as being beneficial to them when they are in close proximity to humans.

Jamel S. Sandidge

Department of Ecology and Evolutionary Biology,
University of Kansas, Lawrence, Kansas 66045, USA
e-mail: spidermn@ku.edu

1. Atkins, J. A., Curtis, W. W., Sodeman, W. W. & Flynn, J. E. Am. J. Trop. Med. Hyg. 7, 165–184 (1958).
2. Gertsch, W. J. & Ennik, F. Bull. Am. Mus. Nat. Hist. 175, 264–360 (1983).
3. Vetter, R. & Barger, D. J. Med. Entomol. 39, 948–951 (2002).
4. Smith, R. B. & Mommson, T. P. Science 226, 1330 (1984).
5. Pollard, S. D. Nat. Hist. 102, 58 (1993).
6. Jackson, R. R. & Blest, A. D. J. Zool. 196, 255–293 (1982).
7. Hite, J. M., Gladney, W. J., Lancaster, J. L. Jr & Whitcomb, W. H. Univ. Arkansas Agric. Exp. Stn Bull. 711, 1–26 (1966).
8. Knost, S. J. & Rovner, J. S. Am. Mid. Nat. 93, 239–244 (1975).

Competing financial interests: declared none.

COMMUNICATIONS ARISING

Biophysics

Is rhodopsin dimeric in native retinal rods?

The model of vertebrate rhodopsin that has emerged from a combination of traditional biophysical techniques^{1–6} defines it as an archetypal monomeric receptor protein that is randomly dispersed and freely diffusing in a fluid lipid matrix in the rod-disc membrane. A quite different arrangement has been demonstrated by Fotiadis *et al.*⁷, however, who use atomic force microscopy (AFM) to reveal rows of rhodopsin molecules packed as dimers in isolated murine disc membranes at twice the surface density estimated from previous *in situ* measurements^{3,4,6}; lipids are probably excluded from these rhodopsin packs⁸. We contend that these 'dimers' are in fact long double rows of equally spaced proteins, packed in partially ordered microcrystalline arrays, and suggest that the arrangement of irregular packs of protein and patches of pure lipid with a free-floating border may be an artefact resulting

from the conditions used in sample preparation for AFM.

Fast, transient, flash-induced photodichroism¹ first revealed the rapid rotational diffusion of rhodopsin *in situ*, and the kinetics of flash-bleaching recovery² indicated that rhodopsin undergoes rapid lateral diffusion in intact rods. Its random distribution in the membrane was evident from X-ray and neutron diffraction^{3–5} and was confirmed by electron microscopy⁶. We therefore question whether the structures seen by Fotiadis *et al.*⁷ in an osmotically burst disc membrane represent a native disc structure, or whether they can be explained by separation of the proteins and lipids during the 15-h incubation at 0 °C in 2 mM Tris–HCl in the preparation of samples for AFM⁸.

A long exposure at low temperature will alter the native disc membrane structure in rods isolated from a warm-blooded vertebrate. In-plane X-ray diffraction revealed long ago that cooling cattle rods to 5 °C for only 15 min induces a partial phase transition of the lipids and a correlated segregation of the proteins⁴. By contrast, no phase separation in frog rods was ever observed by X-ray⁴ or neutron⁵ diffraction.

Low-temperature investigation of flash-induced transient photodichroism (R.C., unpublished results) has confirmed that frog rhodopsin can still undergo rapid rotational diffusion at temperatures close to 0 °C. Transient photodichroism cannot reliably distinguish between freely rotating dimers and monomers, but the packed rows of rhodopsin seen in the AFM images of Fotiadis *et al.*⁷ are incompatible with rapid rotational diffusion. And if rhodopsin were oligomerized into rows *in vivo*, a stable or slowly decaying flash-induced dichroism would have been detected long ago in human vision and in isolated vertebrate retinæ¹.

Electron-microscope images of snap-frozen, freeze-etched frog rods⁶ were able to resolve rhodopsin monomers in a random array, with no evidence of dimers, although the rods had been incubated at 4 °C for 1 h. After digestion of these fragmented rods with phospholipase C, however, disc-membrane phospholipids were seen to segregate into droplets, whereas the packing of rhodopsin changed into concentrated, row-like arrays⁶. It is possible that these could be the equivalent of the double-row arrays that were observed by AFM⁷.

All of these biophysical data are old, but remain valid. Until now, they have not been experimentally challenged or refuted. They exclude the occurrence of rhodopsin oligomers in native, unactivated, frog rods and provide evidence for the rapid phase separation of rhodopsin and lipids in mammalian disc membranes kept at low temperature.

Fotiadis *et al.* observe the same separation of rhodopsin and lipids when their samples are incubated at the higher temperature

of 25 °C (see www.mih.unibas.ch/Nature/Fotiadis.html), but this is still below physiological temperature. Phase separation in a burst disc membrane may slow down under these conditions but will not be prevented. In osmotically disrupted cattle discs, for example, phase-separation artefacts were often visible in early X-ray patterns that required 15 h of exposure at room temperature⁴.

We therefore maintain that native rhodopsin exists in monomeric form in the retinal rods of vertebrates. The double-row arrays seen in AFM^{7,8} in isolated and osmotically disrupted mouse discs probably result from separation of the lipid phase from the proteins, which occurred rapidly at 0 °C and more slowly at 25 °C, during the long incubation and upon osmotic disruption of the discs.

Our contention excludes only constitutive oligomerization of non-activated rhodopsin in dark-adapted rods. After illumination, in an arrestin-mediated deactivation process, rhodopsin molecules may undergo pairing⁹. Oligomerization also occurs in other G-protein-coupled receptors¹⁰ in the β -arrestin-mediated process of their inactivation and clustering in coated pits, but these processes are distinct from the oligomerization of non-activated receptors.

Marc Chabre*, Richard Cone†, Helen Saibil‡

*Institut de Pharmacologie, CNRS-Université de Nice-Sophia Antipolis, 06560 Valbonne, France
e-mail: chabre@ipmc.cnrs.fr

†Department of Biophysics, Johns Hopkins University, Baltimore, Maryland 21218, USA

‡Department of Crystallography, Birkbeck College, London WC1E 7HX, UK

1. Cone, R. A. *Nature New Biol.* **236**, 39–43 (1972).
2. Poo, M. & Cone, R. A. *Nature* **247**, 438–441 (1974).
3. Blasie, J. K., Worthington, C. R. & Dewey, M. M. *J. Mol. Biol.* **39**, 407–416 (1969).
4. Chabre, M. *Biochim. Biophys. Acta* **382**, 322–335 (1975).
5. Saibil, H., Chabre, M. & Worcester, D. L. *Nature* **262**, 266–270 (1976).
6. Roof, D. J. & Heuser, J. E. *J. Cell Biol.* **95**, 487–500 (1982).
7. Fotiadis, D. *et al.* *Nature* **421**, 127–128 (2003).
8. Liang, Y. *et al.* *J. Biol. Chem.* **278**, 21655–21662 (2003).
9. Schröder, K., Pulvermüller, A. & Hofmann, K. P. *J. Biol. Chem.* **277**, 43987–43996 (2002).
10. Bouvier, M. *Nature Rev. Neurosci.* **2**, 274–286 (2001).

Fotiadis et al. reply — Individual biological molecules can be imaged under physiological conditions by atomic force microscopy. Our results from AFM, supported by electron microscopy, revealed distinct rows of rhodopsin dimers and paracrystalline arrays in native murine disc membranes^{1,2}. This supramolecular arrangement was also found for the light-activated form, opsin¹. We counted 30,000–55,000 rhodopsin molecules per square micrometre, a packing density comparable to that measured by optical methods in amphibian discs³. From the lattice vectors describing the paracrystalline arrays, a maximum possible packing density of 62,900 rhodopsin molecules per

square micrometre was calculated. The rhodopsin dimers seen in AFM topographs² have cytosolic protrusions separated by 3.8 nm, providing an ideal docking platform for arrestin, which has two binding grooves that are separated by 3.8 nm as well¹.

On the basis of data reported some 30 years ago, Chabre *et al.* are challenging our observations. Results from their and other³ early experiments, which were all carried out on large ensembles of amphibian disc membranes, indicated that inactivated as well as activated rhodopsin diffuses freely as a monomer in the lipid bilayer. Chabre *et al.* now propose that the rhodopsin-packing arrangement that we observe by AFM^{1,2} is induced by the segregation of proteins and lipids at low temperatures.

To respond to this criticism, we recorded electron micrographs as well as AFM topographs of rhodopsin paracrystals on native disc membranes that had been prepared and imaged at room temperature (see www.mih.unibas.ch/Nature/Fotiadis.html). Three different surface types could again be distinguished: rhodopsin paracrystals, rhodopsin rafts and the lipid bilayer; at increased magnification, rhodopsin dimers, the building block of the paracrystal, are clearly seen. This dimerization and higher-order organization of rhodopsin at room temperature argues against the claim by Chabre *et al.* that the packing arrangement we describe^{1,2} is a result of low-temperature-induced protein–lipid segregation, particularly if we consider that the phase-transition temperature of the bovine-disc lipids is below 3 °C (ref. 4).

Freeze-fracture electron microscopy has also revealed paracrystalline rhodopsin arrays in *Drosophila* photoreceptive membranes⁵ and in the plasma membrane of bovine-rod outer segments⁶. The concept of oligomerization in the presence or absence of ligands is generally accepted for many G-protein-coupled receptors⁷, and rhodopsin is not an exception.

Dimitrios Fotiadis*, Yan Liang†, Slawomir Filipek‡, David A. Saperstein†, Andreas Engel*, Krzysztof Palczewski†§¶

*M.E. Müller Institute for Microscopy, Biozentrum, University of Basel, Basel 4056, Switzerland

Departments of †Ophthalmology, §Pharmacology, and ¶Chemistry, University of Washington, Seattle, Washington 98195, USA

e-mail: palczewski@u.washington.edu

‡International Institute of Molecular and Cell Biology, and Department of Chemistry, University of Warsaw, Warsaw 02109, Poland

1. Liang, Y. *et al.* *J. Biol. Chem.* **278**, 21655–21662 (2003).
2. Fotiadis, D. *et al.* *Nature* **421**, 127–128 (2003).
3. Liebman, P. A. & Entine, G. *Science* **185**, 457–459 (1974).
4. Miljanich, G. P., Brown, M. F., Mabrey-Gaud, S., Dratz, E. A. & Sturtevant, J. M. *J. Membr. Biol.* **85**, 79–86 (1985).
5. Suzuki, E., Katayama, E. & Hirokawa, K. *J. Electron Microsc.* (Tokyo) **42**, 178–184 (1993).
6. Kajimura, N., Harada, Y. & Usukura, J. *J. Electron Microsc.* (Tokyo) **49**, 691–697 (2000).
7. Bouvier, M. *Nature Rev. Neurosci.* **2**, 274–286 (2001).

Metabolic balance of the open sea

The rise of oxygenic photosynthesis nearly three billion years ago led to the accumulation of free oxygen and to the subsequent diversification of life on Earth; today, nearly half of all oxygen production derives from the photosynthetic activities of marine phytoplankton¹. The conclusion that the open sea — and therefore much of our planet's surface — is in a net heterotrophic metabolic state^{2–4} is enigmatic and is a first-order question in the global carbon cycle, as discussed by del Giorgio and Duarte⁵. Our findings suggest that autotrophy in the open sea is episodic and decoupled from the more constant heterotrophic processes. Consequently, the metabolic balance of the open sea depends on proper space and timescale integration to achieve an ecological understanding of life in the sea.

Claims that the open ocean is in a net heterotrophic state run counter to results from other studies from these regions, including net carbon export to deep waters⁶ and net oxygen flux to the atmosphere^{7,8}. How can we resolve this apparent paradox? A recently completed two-year experiment in the oligotrophic North Pacific (near 23° N, 158° W) showed net autotrophy to be variable over time, resulting in short-lived (days to weeks), aperiodic bursts of oxygen accumulation (Fig. 1).

Net heterotrophy is also time-variable, but to a much less extreme extent than net autotrophy, perhaps because of the structure of the microbial community or the dynamics of the separate resource pools. The net

autotrophy events give rise to oxygen supersaturations in the range 2–4% at 50 m (Fig. 1). This is equivalent to a daily net oxygen production of 5–10 mmol m⁻³, about an order of magnitude greater than the 12-year mean production rate based on discrete monthly measurements at this time-series site⁹.

The relatively short lifetime and infrequent occurrence of these events combine to provide a temporal mosaic in the net metabolic state, one that is impossible to assess accurately on the basis of any single sampling (Fig. 1). For example, let us assume that the export ratio is 10%, the background respiration rate is 10% greater than the background rate of oxygen production, and the pulsed rate of oxygen production, which occurs 10% of the time, is three times the background rate: then, if the ecosystem is randomly sampled 1, 5 and 20 times, an observer conducting an oxygen-bottle experiment *in vitro* will incorrectly conclude that the system is net heterotrophic 89%, 55% and 33% of the time, respectively. This probable error in judgement is consistent with the published field results and the general lack of repeat observations. Geochemical techniques, with much longer integration timescales, properly incorporate both net autotrophic and net heterotrophic phases; this is the key to defining the true metabolic state of the sea.

The forcing mechanism(s) responsible for this intermittence in the metabolic state of the sea are not well understood. Although short-term changes in grazing and/or viral lysis could change plant biomass and affect net autotrophy, we currently favour a hypothesis of resource control by one or more novel mechanisms of inorganic nutrient delivery^{10–12} or possibly by the atmospheric deposition of iron¹³. These stochastic

nutrient-loading processes can rapidly alter microbial community structure, decouple organic matter cycles, and lead to the high-frequency net autotrophy events at station ALOHA (Fig. 1). At present, no ocean observatory has an adequate measurement programme in place to detect these proposed events and therefore to assess the metabolic balance of the open sea. This is an important challenge for the future because accurate estimation and comprehensive understanding of the metabolic balance of the open sea have implications for the global carbon cycle and carbon sequestration processes.

David M. Karl*, Edward A. Laws*, Paul Morris*, Peter J. leB. Williams†, Steven Emerson‡

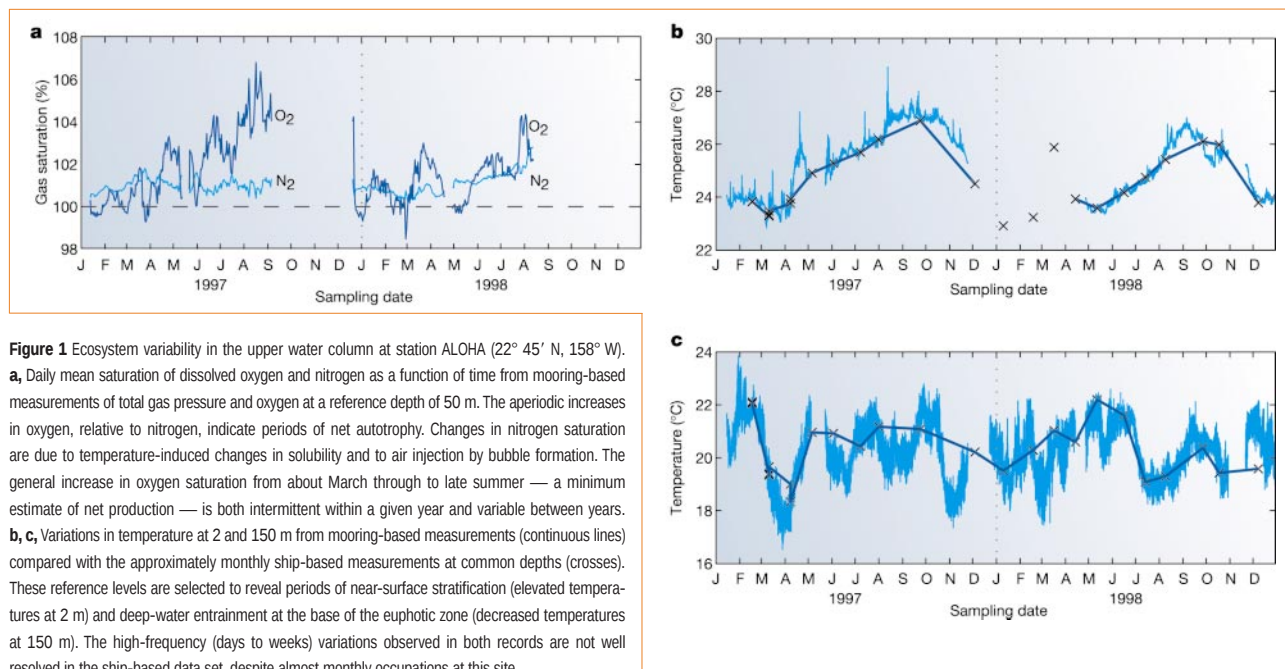
*Department of Oceanography, University of Hawaii, Honolulu, Hawaii 96822, USA

e-mail: dkarl@hawaii.edu

†School of Ocean Sciences, University of Wales, Bangor LL59 5PP, UK

‡School of Oceanography, University of Washington, Seattle, Washington 98195, USA

1. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. *Science* **281**, 237–240 (1998).
2. del Giorgio, P. A., Cole, J. J. & Cimbleris, A. *Nature* **385**, 148–151 (1997).
3. Duarte, C. M. & Agusti, S. *Science* **281**, 234–236 (1998).
4. Duarte, C. M., Agusti, S., del Giorgio, P. A. & Cole, J. J. *Science* **284**, 1735 (1999).
5. del Giorgio, P. A. & Duarte, C. M. *Nature* **420**, 379–384 (2002).
6. Emerson, S. *et al. Nature* **389**, 951–954 (1997).
7. Emerson, S., Quay, P. D., Stump, C., Wilbur, D. & Schudlich, R. *J. Geophys. Res.* **100**, 15873–15887 (1995).
8. Najjar, R. G. & Keeling, R. F. *Global Biogeochem. Cycles* **14**, 573–584 (2000).
9. Karl, D. M., Bidigare, R. R. & Letelier, R. M. in *Phytoplankton Productivity and Carbon Assimilation in Marine and Freshwater Ecosystems* (eds Williams, P. J. leB., Thomas, D. R. & Reynolds, C. S.) 222–264 (Blackwell, London, 2002).
10. McGillicuddy, D. J. Jr *et al. Nature* **394**, 263–266 (1998).
11. Uz, B. M., Yoder, J. A. & Osychyn, V. *Nature* **409**, 597–600 (2001).
12. Gregg, M. C., Sanford, T. B. & Winkel, D. P. *Nature* **422**, 513–515 (2003).
13. Young, R. W. *et al. Global Biogeochem. Cycles* **5**, 119–134 (1991).



The immune response of *Drosophila*

Jules A. Hoffmann

Institut de Biologie Moléculaire et Cellulaire du CNRS, 67084 Strasbourg Cedex, France

***Drosophila* mounts a potent host defence when challenged by various microorganisms. Analysis of this defence by molecular genetics has now provided a global picture of the mechanisms by which this insect senses infection, discriminates between various classes of microorganisms and induces the production of effector molecules, among which antimicrobial peptides are prominent. An unexpected result of these studies was the discovery that most of the genes involved in the *Drosophila* host defence are homologous or very similar to genes implicated in mammalian innate immune defences. Recent progress in research on *Drosophila* immune defence provides evidence for similarities and differences between *Drosophila* immune responses and mammalian innate immunity.**

Metazoans have developed efficient mechanisms to oppose microbial invaders. Innate immunity is common to all metazoans and serves as a first-line defence. Its hallmarks are the recognition of microorganisms by germline-encoded, non-rearranging receptors, and rapid effector mechanisms that involve phagocytosis, activation of proteolytic cascades and synthesis of potent antimicrobial peptides^{1,2}. Adaptive immunity appeared in the ancestors of cartilaginous fish and is restricted today to some 45,000 vertebrate species, where it coexists with innate immune defences. Adaptive immunity depends on the generation of a complex repertoire of immune receptors in lymphocytes through somatic gene rearrangement, and on the clonal expansion of activated lymphocytes. In contrast to innate immunity, adaptive immunity is endowed with memory. Being devoid of adaptive immune reactions, the *Drosophila* host defence seems to be a model system for the study of pristine innate immune defences. The antimicrobial response in this species, as in other metazoans, comprises both cellular and humoral reactions. The cellular defence is best illustrated by the strong phagocytic activities of the predominant blood cells, the plasmatocytes^{3,4}. The hallmark of the humoral reactions is the induction by microbial challenge of antimicrobial peptide genes in the fat body (equivalent of the mammalian liver), followed by the secretion of these peptides into the haemolymph (blood)⁵. We will focus on this aspect in the present review, keeping in mind that other humoral reactions—for example, melanization—also contribute to the *Drosophila* host defence⁴. Furthermore, as in other animal phyla, epithelial immune reactions, which involve the local synthesis of antimicrobial peptides, are activated in response to infection^{6,7}.

Seven distinct *Drosophila* inducible antimicrobial peptides (or peptide families) have now been identified⁸. They are structurally diverse and, as their mammalian counterparts, are mostly small in size (5 kDa), cationic and predominantly membrane active. Their activity spectra are directed either against fungi (Drosomycins, Metchnikowin), or against Gram-positive (Defensin) and Gram-negative (Attacins, Cecropins, Drosocin, Dipterocins) bacteria. It is assumed that their combined activities largely contribute to blocking the growth of invading microorganisms in the haemolymph.

The promoter regions of the antimicrobial peptide genes contain nucleotide motifs similar to mammalian binding sites for NF- κ B/Rel proteins. Mutations of these sequences in the early 1990s abolished the immune-inducibility of the corresponding genes in reporter fly lines^{9,10}. These findings raised the question of whether an equivalent of an NF- κ B-based transcriptional activation mechanism was involved in the *Drosophila* host defence. *Drosophila* was known, at that time, to express a gene that encodes a member of the NF- κ B family. This gene, *dorsal*, had been genetically identified as a regulator of dorsoventral patterning in the early embryo¹¹. It was also known that in the embryo, dissociation of the Dorsal protein

from its inhibitor, the Cactus protein, was dependent on activation of the transmembrane receptor Toll by a proteolytically cleaved form of a cytokine-like protein, Spaetzle (reviewed in refs 12, 13).

The tantalizing similarities between the activation of Dorsal in the *Drosophila* embryo and the cytokine-induced activation of NF- κ B during inflammatory responses in mammals (reviewed in ref. 14) prompted our group to undertake an in-depth genetic analysis of the immune defences in flies mutant for the various genes of the dorsoventral regulatory cascade. A striking result, published in 1996, was that loss-of-function mutations in the Toll receptor compromised the survival of flies faced with fungal infection and the challenge-dependent transcription of the antifungal peptide Drosomycin¹⁵. These data provided the first example of an immune function for a transmembrane receptor that combined an extracellular leucine-rich domain and an intracytoplasmic TIR (for Toll/interleukin-1 receptor) homology domain. At that time, it had been shown that bacterial lipopolysaccharide (LPS) binds on the membranes of immune-responsive mammalian cells to CD14, a glycosyl phosphatidylinositol (GPI)-anchored protein with leucine-rich repeat domains (as in Toll)¹⁶. However, as CD14 by itself lacks the capability to transmit an intracellular signal to activate NF- κ B, the question of how LPS activated NF- κ B remained open. The finding that *Drosophila* Toll had an immune function prompted the search for Toll-like receptors in mammals. In 1997, the successful cloning of a human homologue of *Drosophila* Toll was reported by Janeway and Medzhitov, and this Toll-like receptor (TLR) was shown to activate NF- κ B¹⁷. The immune relevance of this TLR was dramatically highlighted when Beutler and associates discovered that the *lps* mutation in mice, which abolishes the response to bacterial LPS, corresponded to a loss-of-function mutation in this receptor^{18,19}. These results, and the concomitant discovery that mammalian genomes encoded a family of TLRs²⁰, opened a new field in the study of recognition of infectious non-self, and more generally rejuvenated interest in mammalian innate immune reactions.

Coming back to *Drosophila*, the genetic analyses of the *Drosophila* host defence mentioned above had indicated that activation of the Toll pathway accounted primarily for the response to infections by fungi and Gram-positive bacteria²¹. A mutation in a gene named *imd* (for immune deficiency) was found to affect rather more the resistance of flies to Gram-negative bacterial sepsis²². Consequently, the *imd* gene was proposed to define a second immune signalling pathway. The progress made recently in our understanding of the roles of the Toll and Imd pathways in *Drosophila* immunity is outlined below.

Toll(s) in the host defence of *Drosophila*

Toll activation during the immune response (Fig. 1) is strictly dependent on the product of the *Spaetzle* gene. The Spaetzle protein is a cystine-knot molecule with structural similarities to mamma-

lian neurotrophins, and requires proteolytic cleavage for full biological activity^{23,24}. This cleavage is induced by a proteolytic cascade activated as an early result of infection. The mature 12-kDa form of Spaetzle binds as a dimer to the Toll ectodomain with high affinity ($K_d \approx 0.4$ nM) and with a stoichiometry of one Spaetzle dimer to two receptor proteins²⁵. The intracytoplasmic TIR domain of Toll interacts with three partners, each of which has a death domain. Two of these are adaptor proteins: the *Drosophila* homologue of MyD88^{26–29}, which in addition to the death domain has a TIR domain similar to that of Toll with which it associates, and Tube. Tube has no obvious mammalian homologue. The third death-domain protein in this receptor–adaptor complex is Pelle, which has a serine–threonine kinase domain and is homologous to mammalian IRAKs (interleukin-1 receptor-associated kinases; reviewed in ref. 30). Depending on the developmental stage, Toll can activate two closely related NF- κ B proteins in immune-responsive tissues: DIF³¹ (Dorsal-related immunity factor) in adults, and Dorsal and/or DIF in larvae^{32–34}. The end effect of Toll signalling is the dissociation of NF- κ B protein from the ankyrin-repeat inhibitory protein Cactus, a homologue of mammalian I κ Bs. This process involves signal-dependent phosphorylation of Cactus, followed by its degradation by the proteasome^{35,36}. The activation of Dorsal requires phosphorylation, in addition to dissociation from Cactus (see also^{37,38}). It is unclear how activation of the Toll receptor–adaptor complex leads to these various processes. Although

Drosophila expresses genes encoding members of the TRAF (TNF-receptor-associated factor) family and homologues of mammalian IKK- β (I κ B kinase- β) and IKK- γ /NEMO, genetic studies have failed so far to demonstrate an involvement of any of these genes downstream of Toll. Furthermore, Pelle does not directly phosphorylate Cactus and the identity of the Cactus kinase remains elusive.

The precise roles of the Toll pathway during the response to fungal and Gram-positive bacterial infection are not fully understood. One effect is obviously to direct the expression of various antimicrobial peptides. However, microarray data have indicated that hundreds of genes are markedly upregulated as a consequence of the challenge-dependent activation of Toll^{39,40}, and their functions have not yet been adequately addressed. In addition to Toll, the *Drosophila* genome contains eight homologues (18-Wheeler/Toll-2 to Toll-9)⁴¹. Except for Toll, it has not been possible to unequivocally attribute an immune function to any of the members of the family. In contrast to mammalian TLRs, expression of the *Drosophila* Tolls is not restricted to immune-responsive cells. Rather, Tolls show complex stage- and tissue-specific expression patterns during embryonic and larval development⁴². *Drosophila* also expresses six distinct Spaetzle-like genes, particularly during embryonic development⁴³. The various Spaetzle proteins are obvious candidate ligands for the members of the Toll family. In addition, homophilic and/or heterophilic interactions between leucine-rich repeat ectodomains of Tolls of neighbouring cells during development have been suggested as a possible mode of activation⁴⁴. In summary, the Toll receptors seem to have specific developmental roles, with Toll being, in addition, recruited for a defence function.

The Imd pathway

The Imd pathway (Fig. 1) is primarily, although not exclusively, involved in defence against Gram-negative bacteria, and most of the antimicrobial peptides controlled by this pathway (Diptericin, Drosocin, Cecropins, Attacins) have activity spectra predominantly directed against Gram-negative bacteria. The *imd* gene encodes a 25-kDa protein with a death domain that has strongest similarities to that of mammalian RIP (TNF-receptor-interacting protein)⁴⁵. However, in contrast to RIP, the Imd protein has no kinase domain. The NF- κ B protein involved in the Imd pathway, Relish, is not inhibited by Cactus, but carries its own inhibitory ankyrin-repeat sequences in its carboxy-terminal region^{46,47}. Activation of Relish therefore requires a signal-induced endoproteolytic cleavage that frees the Rel-homology domain from the inhibitory domain and allows its nuclear translocation⁴⁸. The sequence of events downstream of Imd that lead to phosphorylation and ultimately to cleavage of Relish are not fully deciphered. However, genetic data point to the involvement of several players, which all have mammalian homologues: (1) *Drosophila* FADD, which associates with Imd^{49,50}; (2) a mitogen-activated protein (MAP) kinase kinase homologous to TAK1 (TGF- β -activated kinase 1)⁵¹, which acts downstream of Imd/FADD; (3) a *Drosophila* equivalent of the mammalian signalosome comprising IKK- β and IKK- γ /NEMO homologues^{52–54}. This last complex is probably activated by TAK1 and directs phosphorylation of Relish, which precedes its proteolytic cleavage⁵⁴. Mutations in any of these genes considerably decrease the immune-inducibility of antibacterial peptide genes and compromise survival against Gram-negative bacterial strains. Furthermore, the *Drosophila* caspase-8 homologue DREDD is required for full activation of Relish^{48,55,56}. DREDD can associate with FADD and with Relish, and it has been proposed recently that this caspase directly cleaves Relish^{57,58}.

Of significant interest is the recent demonstration, based on a comprehensive microarray analysis in *Drosophila* cell lines, that the Imd pathway can branch downstream of TAK1 and, in addition to controlling antimicrobial peptide synthesis through Relish, activates the expression of cytoskeletal proteins through a JNK (c-Jun N-terminal kinase) cascade. As proposed by Perrimon and associ-

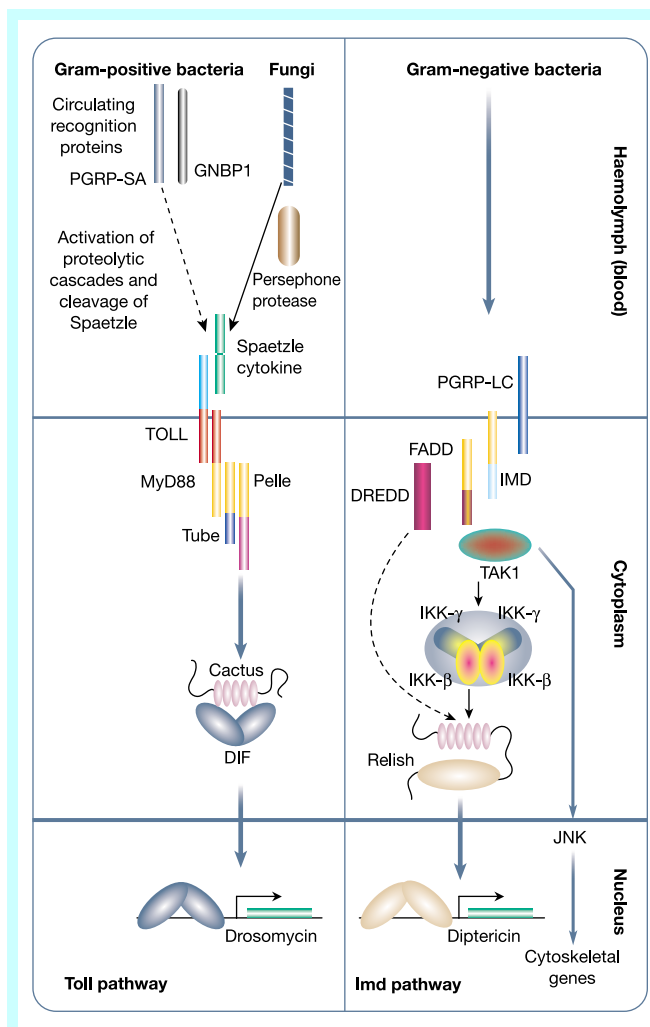


Figure 1 The Toll and Imd pathways in the control of expression of genes encoding antimicrobial peptides. For explanations, see the main text. Colour codes for common domains: yellow, death domains; red, TIR domains.

ates⁵⁹, co-regulation in response to immune challenge of JNK and NF- κ B (Relish) through the same upstream intracellular pathway could allow a close linkage between antimicrobial defences and tissue repair processes. As in the case of Toll, microarray analyses have shown that the Imd pathway controls the transcription of several hundred genes in addition to that of the antibacterial peptide genes. The functions of most of these genes in host defence remain to be established^{40,60}.

Activation of Toll and Imd signalling pathways

As indicated above, Toll is activated by interaction with a cleaved form of the cytokine-like polypeptide Spaetzle, and not by direct interaction with microbial inducers. How are these inducers recognized during infection? A genetic analysis has recently led to the identification of two genes—*semmelweis* and *osiris*—that are required for Toll activation by Gram-positive bacterial infection^{61,62}. Each gene encodes a member of a family of proteins, the PGRPs (peptidoglycan-recognition proteins) and the GNBPs (Gram-negative-binding proteins), that were initially characterized in larger insect species through their ability to bind either to intact bacteria or to structural motifs of bacteria^{63–66} (Fig. 2). The *Drosophila* genome contains 13 genes belonging to the PGRP family and three genes of the GGBP family. Seven of the *Drosophila* PGRPs are small-sized (20 kDa) extracellular polypeptides, whereas other members are larger (30 to 90 kDa) and are predicted to be either intracellular or membrane spanning. Some of the genes encoding the larger PGRPs form splice isoforms, bringing the potential number of PGRP forms in this species close to 20 (ref. 67). GNBPs are 50-kDa proteins with a C-terminal, 200-residue β -glucanase-like domain and an N-terminal, 100-residue domain reported to bind β -(1,3)-glucan⁶⁸. Intriguingly, loss-of-function mutations in either *semmelweis* (PGRP-SA) or *osiris* (GNBP1) give very similar phenotypes of compromised survival to Gram-positive bacterial infection and activation of Toll. These two blood-borne proteins probably cooperate to sense Gram-positive bacteria and to activate proteolytic enzymes(s) leading to the cleavage of Spaetzle. This view is substantiated by the observation that concomitant overexpression of the *semmelweis* and the *osiris* genes results in the challenge-independent induction of the Toll pathway⁶². Activation of Toll by fungal infections is independent of the *semmelweis* and *osiris* genes. A genetic screen has now identified a gene (*persephone*) encoding a haemolymph trypsin-like protease that mediates the fungal-dependent cleavage of Spaetzle and activation of Toll. Overexpression of the *persephone* gene is sufficient to lead to Spaetzle-dependent induction of the Toll pathway in the absence of an immune challenge⁶⁹. The recognition proteins for fungal infection and the proteases activated downstream of Semmelweis (PGRP-SA)/Osiris (GNBP1) have not yet been identified.

Activation of the Imd pathway requires a putative transmembrane protein that is also a member of the PGRP family (PGRP-LC)^{70–72}. The molecular mechanisms by which PGRP-LC activates the Imd pathway remain to be worked out: the amino-acid sequence of the putative intracytoplasmic domain of this transmembrane protein is not informative in this respect. As loss-of-function mutants of PGRP-LC have less severe phenotypes than null mutants of *imd*, it is probable that PGRP-LC is only one of several members of a complex that senses Gram-negative bacteria. Overexpression of yet another member of the PGRP family, PGRP-LE, can also activate the Imd pathway⁷³.

The data presented above underline the paramount importance of PGRP family members in sensing microbial infections. PGRPs have in common one or several domains of 160 amino acid residues. A structure-based sequence alignment has identified conserved features between PGRP domains and bacterial amidases, which cleave peptidoglycans^{74,75}. Sequence changes in these domains indicate that some of the PGRPs have lost their amidase function and serve essentially for recognition, whereas others, such as PGRP-SC1, have retained their enzymatic activity⁷⁶. The crystal structure of one of the PGRPs has now also been reported and points to an active-site cleft with a zinc cage⁷⁷. The differences in structures of the PGRP domains of distinct PGRPs may account for the discrimination between classes of microorganisms. Gram-positive lysine-peptidoglycan has recently been shown to activate the Toll pathway in a PGRP-SA (Semmelweis)-dependent manner, whereas Gram-negative diaminopimelic acid peptidoglycan induces the Imd pathway through PGRP-LC⁷⁸. Bacilli, which are Gram-positive bacteria that synthesize diaminopimelic acid peptidoglycan, activate the Imd pathway through this molecular pattern. They also activate the Toll pathway, probably through distinct microbial patterns. This example cautions against establishing too strict a link between Gram-positive infection and induction of the Toll pathway, versus Gram-negative infection and activation of the Imd pathway.

The discovery of insect PGRPs and their immune functions prompted the search for mammalian homologues, and it is now established that mice and humans express four genes encoding members of this family⁷⁹. One of these is a small form, PGRP-S, present in neutrophil tertiary granules. A second PGRP is larger and membrane spanning, and two further forms are of intermediate size. Mice in which the PGRP-S gene has been knocked out show impaired intracellular killing of low-pathogenicity Gram-positive bacteria⁸⁰. It is premature to assess the contribution of PGRPs to the overall host defence in mammals, where peptidoglycan is essentially sensed by TLR2 in immune-responsive cells⁸¹ (and by NODs in epithelial cells^{82,83}). By contrast, in view of the genetic data, we can be confident that, in *Drosophila*, PGRPs have a paramount,

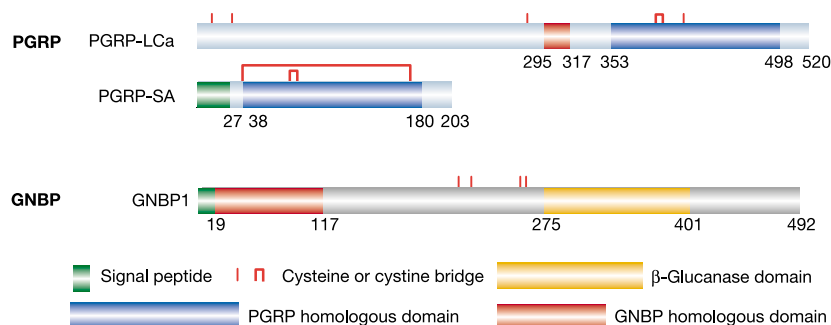


Figure 2 Selected members of the PGRP and GGBP families. The large membrane-spanning form PGRP-LC is involved in activation of the Imd pathway (here, one of several isoforms, LCa, is presented), and the small circulating PGRP-SA (Semmelweis)

is required for Toll activation by Gram-positive bacteria, together with GNBP1 (Osiris). For details, see main text.

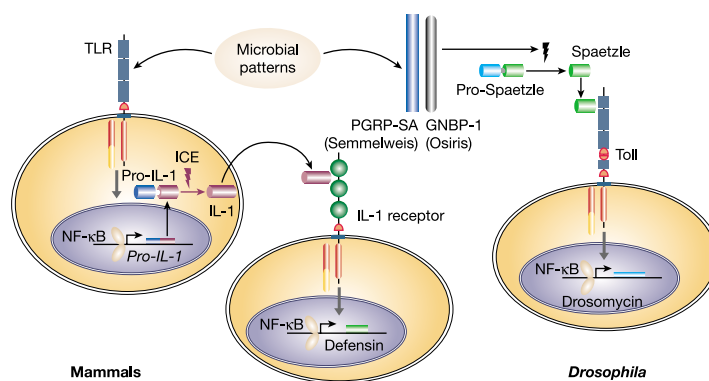


Figure 3 Parallels between the functions of interleukin-1 and Spaetzle. Recognition of Gram-positive bacterial infection is mediated by TLRs on immune-responsive cells in mammals, and by circulating pattern-recognition proteins in *Drosophila*. The part of the

figure devoted to mammals illustrates the induction of human β -defensin-2 in keratinocytes under the stimulus of interleukin-1 (IL-1) produced by macrophages in response to LPS in the epidermis. See text for further comments.

although not unique, role in the recognition of infection that seems to parallel that of mammalian TLRs.

Mammalian innate immune signalling

Early results on the roles of Toll and TLRs in fighting microbial infections generated an enthusiastic view that we were in the presence of a primeval defence mechanism that had evolved before the separation between the ancestors of present-day insects and mammals. This view was rapidly challenged by a series of observations, namely (1) that Toll does not directly sense microbial inducers, in stark contrast to TLRs, and (2) that two distinct signal-transduction pathways are required in *Drosophila* to respond to Gram-positive and Gram-negative infections, whereas a single TLR-dependent pathway is sufficient in mammals. The overall picture, which has evolved over the past few years, now points both to stringent similarities and to some marked differences. For one, a common thread in both systems is that infection leads to activation of NF- κ B family members and subsequent expression of immune-response genes. This is obviously not the only defence mechanism, but it is a central theme of insect and mammalian host responses. As pointed out above, NF- κ B activation in *Drosophila* is either Toll dependent (dissociation of DIF from Cactus and nuclear translocation of DIF) or occurs through Imd (proteolytic cleavage of Relish and nuclear translocation of the N-terminal Rel domain), depending on the nature of the microbial inducer. The Toll-dependent activation of the NF- κ B family member DIF, as far as we understand it today, relies on intracellular partners with clear homology to TLR partners, namely MyD88 and Pelle/IRAK. The case is often made that Tube is functionally equivalent to TLR adaptor proteins. Although *Drosophila* has homologues of mammalian IKK- β and IKK- γ , genetic evidence indicates that they do not have a role in the Toll pathway, but rather in the Imd pathway. Activation of Toll is obviously different from that of TLRs, as interaction occurs with a cleaved cytokine in one case, as opposed to a direct interaction with microbial inducers in the other case. It is of interest here to note that, in mammals, TLR-dependent expression of antimicrobial peptides can also depend on an activated cytokine, in addition to being directly inducible by TLRs⁸⁴. Experiments reported recently by Ganz and colleagues⁸⁵ are particularly informative in this respect. In the human epidermis, Langerhans cells, monocytes and macrophages respond to bacterial LPS by means of TLR4, and produce significant amounts of interleukin-1. In turn, this cytokine induces (after its proteolytic cleavage to the mature form) the massive and selective synthesis of human β -defensin-2 (HBD-2) in neighbouring keratinocytes. Figure 3 compares these data with the immune induction of Drosomycin in *Drosophila*, where

microbial patterns (for example, peptidoglycan) are sensed by circulating proteins (for example, PGRPs), leading to cleavage of the cytokine Spaetzle and activation of Toll with the subsequent induction of the *Drosomycin* gene. In both cases, expression of the effector molecules (HBD-2, respectively Drosomycin) is directed by a cytokine (interleukin-1, Spaetzle). The synthesis or activation of the cytokines is induced when microbial patterns interact either with TLRs on immune-responsive cells, or with PGRPs in the circulating haemolymph. The inherent anatomical differences between the mammalian and insect systems might explain why recognition and amplification occurs directly on immune-responsive cells in one case and in circulation in the other. Again, there are differences here, namely that the processing of pro-interleukin-1 and that of pro-Spaetzle occur through classes of proteases (caspase versus serine protease) that are not evolutionarily related. However the precise examples presented in Fig. 3 illustrate the similarities between some of the strategies used both by insects and mammals to sense infection and amplify the information.

As regards the Imd-dependent activation of the NF- κ B family member Relish, it came as a surprise to find that several molecules within the signalling pathway were either homologous or very similar to members of pathways initiated by tumour-necrosis factor receptors (TNFRs) in mammals (FADD; DREDD/caspase-8, the death domains of Imd and RIP) (Fig. 4). Furthermore, Relish activation is dependent on homologues of TAK1 and of IKK- β and IKK- γ . At this moment, there is no conclusive evidence for the involvement of *Drosophila* TRAF homologues in this pathway (nor in the Toll pathway). We have been unable to show that *Drosophila* uses apoptosis as a defence mechanism, although we have observed that overexpression of the *Imd* gene induces apoptosis in the fat-body cells⁴⁵. Whereas activation of the Imd pathway by microbial patterns can probably be directly mediated by the PGRP-LC transmembrane receptor (and associated proteins), NF- κ B activation through the members of the TNFR pathway mentioned above cannot occur through direct sensing of microbes by these receptors. In this case, it is the interaction of microbial patterns with TLRs that induces an NF- κ B-dependent synthesis of TNF, which, upon release (autocrine, paracrine), activates the cognate receptor(s).

In conclusion, *Drosophila* NF- κ B activation in response to an immune challenge relies on two distinct pathways, each of which comprises molecules that are homologous or related to molecules found in mammalian signalling pathways activated during innate immune defences. Although neither of the two pathways is *sensu stricto* equivalent to any of those in mammals, it is obvious that most players of the Toll and Imd pathway are also found in either the TLR or/and the TNFR pathways, although not necessarily with the same

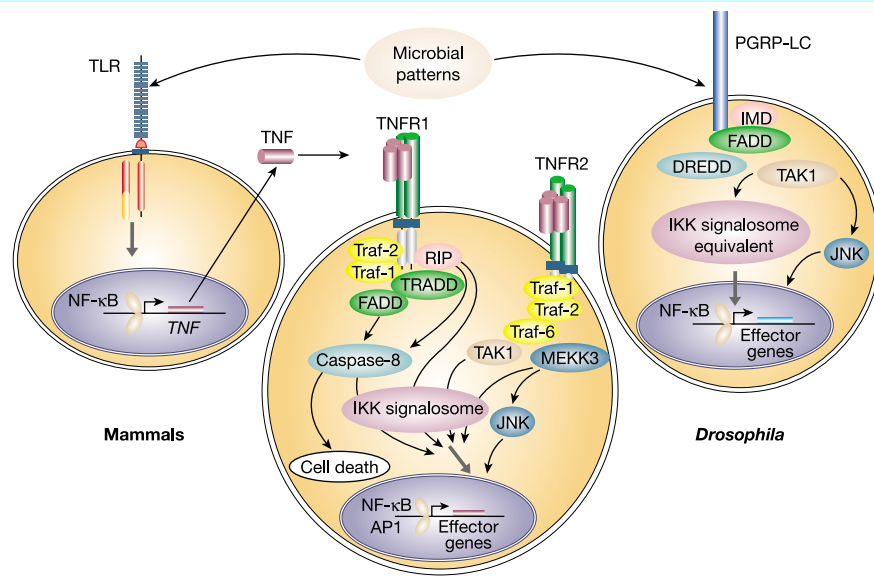


Figure 4 Parallels between the TNFR and Imd pathways. Recognition of Gram-negative bacterial infection in *Drosophila* is mediated, in part, by the putative transmembrane receptor PGRP-LC, and activates NF-κB through a signal-transduction pathway that

involves several molecules homologous or related to signalling partners acting downstream of TNFRs. Similarities and differences are discussed in the main text.

precise functions. The prevailing impression is that the *Drosophila* and mammalian immune pathways have evolved from a reduced number of common ancestral building blocks to the present configuration. Whether some of the stringent parallels, on which we tend to focus, result from convergent evolution or reflect common ancestral pathways is impossible to decide. Molecular genetic analysis of immune defence in lower eukaryotes will probably be revealing in this respect.

The golden apples of the Hesperides

The study of immune defences in *Drosophila* is a relatively junior field of research. The harvest in this field has been rich in recent years, as illustrated above. Furthermore, *Drosophila* can be credited with having largely contributed to the renewed interest in innate immunity. In particular, the exploding field of Toll-like receptors has benefited substantially from the studies of antimicrobial defences in the fly. This is an appropriate moment to reflect on where we anticipate the field to go within the next decade. To borrow an expression from Greek mythology: where are the golden apples of the Hesperides in the field?

A first golden apple will be the molecular definition of the mechanisms that link the recognition of invading microorganisms to activation of the Toll and Imd pathways. This will require the full identification of the various ligands that are recognized by distinct *Drosophila* recognition proteins, and analysis of the structural and functional changes that result from their binding to confer activating properties on these proteins. The observation that two distinct types of blood protein cooperate in the recognition of Gram-positive bacteria illustrates that it will not be trivial to pick this apple.

A second golden apple will be an understanding of the specificity of the host response to distinct classes of microbial invaders. Although the term 'specificity' may be an overstatement in this context, there is clearly a difference in response with regard to the nature of the aggressor. A first line of discrimination lies obviously with the preferential activation of Toll or Imd by different types of microorganism. However, this is not sufficient to account for the different patterns of responses. To give but one example of these differences, fungi and Gram-positive bacteria both activate the Toll pathway, yet microarray analysis shows that only one-third of the genes induced by these infections are common.

A third apple will be gaining a comprehensive view of the effector mechanisms that concur to eliminate infection in *Drosophila*. Research in this area would obviously be helped by a better knowledge of natural diseases that plague these insects in the field. At present, we see resistance to infection as a combined effect of phagocytosis (and encapsulation) by blood cells and massive production of antimicrobial peptides. Phenoloxidase cascades at wound sites may contribute by generating toxic quinones. But this view is superficial. The precise roles of the antimicrobial peptides are not fully understood. To add to the list of unknowns, half of the hundreds of genes significantly upregulated after various microbial challenges encode proteins with unknown functions. The comprehensive view of elimination of infection will undoubtedly serve as a meaningful basis for the study of host–pathogen interaction, which by itself represents one of the most valuable goals in the field. □

doi:10.1038/nature02021.

- Medzhitov, R. & Janeway, C. Jr Innate immunity. *N. Engl. J. Med.* **343**, 338–344 (2000).
- Janeway, C. A. Jr & Medzhitov, R. Innate immune recognition. *Annu. Rev. Immunol.* **20**, 197–216 (2002).
- Rizki, R. M. & Rizki, T. M. Selective destruction of a host blood cell type by a parasitoid wasp. *Proc. Natl Acad. Sci. USA* **81**, 6154–6158 (1984).
- Braun, A., Hoffmann, J. A. & Meister, M. Analysis of the *Drosophila* host defense in *domino* mutant larvae, which are devoid of hemocytes. *Proc. Natl Acad. Sci. USA* **95**, 14337–14342 (1998).
- Hoffmann, J. A. & Reichhart, J. M. *Drosophila* innate immunity: an evolutionary perspective. *Nature Immunol.* **3**, 121–126 (2002).
- Ferrandon, D. et al. A *drosomycin*-GFP reporter transgene reveals a local immune response in *Drosophila* that is not dependent on the Toll pathway. *EMBO J.* **17**, 1217–1227 (1998).
- Tzou, P. et al. Tissue-specific inducible expression of antimicrobial peptide genes in *Drosophila* surface epithelia. *Immunity* **13**, 737–748 (2000).
- Bulet, P., Hetru, C., Dimarcq, J. L. & Hoffmann, D. Antimicrobial peptides in insects: structure and function. *Dev. Comp. Immunol.* **23**, 329–344 (1999).
- Engstrom, Y. et al. κB-like motifs regulate the induction of immune genes in *Drosophila*. *J. Mol. Biol.* **232**, 327–333 (1993).
- Kappler, C. et al. Insect immunity. Two 17 bp repeats nesting a κB-related sequence confer inducibility to the dipterin gene and bind a polypeptide in bacteria-challenged *Drosophila*. *EMBO J.* **12**, 1561–1568 (1993).
- Nusslein-Volhard, C., Lohs-Schardin, M., Sander, K. & Cremer, C. A dorso-ventral shift of embryonic primordia in a new maternal-effect mutant of *Drosophila*. *Nature* **283**, 474–476 (1980).
- St Johnston, D. & Nusslein-Volhard, C. The origin of pattern and polarity in the *Drosophila* embryo. *Cell* **68**, 201–219 (1992).
- Morisato, D. & Anderson, K. V. Signaling pathways that establish the dorsal–ventral pattern of the *Drosophila* embryo. *Annu. Rev. Genet.* **29**, 371–399 (1995).
- Belvin, M. P. & Anderson, K. V. A conserved signaling pathway: the *Drosophila* Toll–Dorsal pathway. *Annu. Rev. Cell Dev. Biol.* **12**, 393–416 (1996).
- Lemaître, B., Nicolas, E., Michaut, L., Reichhart, J. M. & Hoffmann, J. A. The dorsoventral regulatory gene cassette *spatzle*/Toll/cactus controls the potent antifungal response in *Drosophila* adults. *Cell* **86**, 973–983 (1996).

16. Tobias, P. S., Soldau, K. & Ulevitch, R. J. Identification of a lipid A binding site in the acute phase reactant lipopolysaccharide binding protein. *J. Biol. Chem.* **264**, 10867–10871 (1989).
17. Medzhitov, R., Preston-Hurlburt, P. & Janeway, C. A. Jr A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* **388**, 394–397 (1997).
18. Poltorak, A. *et al.* Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in *Tlr4* gene. *Science* **282**, 2085–2088 (1998).
19. Qureshi, S. T. *et al.* Endotoxin-tolerant mice have mutations in Toll-like receptor 4 (Tlr4). *J. Exp. Med.* **189**, 615–625 (1999).
20. Rock, F. L., Hardiman, G., Timans, J. C., Kastelein, R. A. & Bazan, J. F. A family of human receptors structurally related to *Drosophila* Toll. *Proc. Natl Acad. Sci. USA* **95**, 588–593 (1998).
21. Rutschmann, S., Kilinc, A. & Ferrandon, D. Cutting edge: the Toll pathway is required for resistance to Gram-positive bacterial infections in *Drosophila*. *J. Immunol.* **168**, 1542–1546 (2002).
22. Lemaitre, B. *et al.* A recessive mutation, immune deficiency (*imd*), defines two distinct control pathways in the *Drosophila* host defense. *Proc. Natl Acad. Sci. USA* **92**, 9465–9469 (1995).
23. Mizuguchi, K., Parker, J. S., Blundell, T. L. & Gay, N. J. Getting knotted: a model for the structure and activation of Spätzle. *Trends Biochem. Sci.* **23**, 239–242 (1998).
24. DeLotto, Y. & DeLotto, R. Proteolytic processing of the *Drosophila* Spätzle protein by easter generates a dimeric NGF-like molecule with ventralising activity. *Mech. Dev.* **72**, 141–148 (1998).
25. Weber, A. N. R. *et al.* Binding of the *Drosophila* cytokine Spaetzle to Toll is direct and establishes signaling. *Nature Immunol.* **8**, 794–800 (2003).
26. Horng, T. & Medzhitov, R. *Drosophila* MyD88 is an adapter in the Toll signaling pathway. *Proc. Natl Acad. Sci. USA* **98**, 12654–12658 (2001).
27. Sun, H., Bristow, B. N., Qu, G. & Wasserman, S. A. A heterotrimeric death domain complex in Toll signaling. *Proc. Natl Acad. Sci. USA* **99**, 12871–12876 (2002).
28. Tauszig-Delamasure, S., Bilak, H., Capovilla, M., Hoffmann, J. A. & Imler, J. L. *Drosophila* MyD88 is required for the response to fungal and Gram-positive bacterial infections. *Nature Immunol.* **3**, 91–97 (2002).
29. Charatsi, I., Luschni, S., Bartoszewski, S., Nusslein-Volhard, C. & Moussian, B. Krapfen/dMyd88 is required for the establishment of dorsoventral pattern in the *Drosophila* embryo. *Mech. Dev.* **120**, 219–226 (2003).
30. Janssens, S. & Beyaert, R. Functional diversity and regulation of different interleukin-1 receptor-associated kinase (IRAK) family members. *Mol. Cell* **11**, 293–302 (2003).
31. Ip, Y. T. *et al.* Dif, a dorsal-related gene that mediates an immune response in *Drosophila*. *Cell* **75**, 753–763 (1993).
32. Meng, X., Khanuja, B. S. & Ip, Y. T. Toll receptor-mediated *Drosophila* immune response requires Dif, an NF- κ B factor. *Genes Dev.* **13**, 792–797 (1999).
33. Manfrulli, P., Reichhart, J. M., Steward, R., Hoffmann, J. A. & Lemaitre, B. A mosaic analysis in *Drosophila* fat body cells of the control of antimicrobial peptide genes by the Rel proteins Dorsal and DIF. *EMBO J.* **18**, 3380–3391 (1999).
34. Rutschmann, S. *et al.* The Rel protein DIF mediates the antifungal but not the antibacterial host defense in *Drosophila*. *Immunity* **12**, 569–580 (2000).
35. Nicolas, E., Reichhart, J. M., Hoffmann, J. A. & Lemaitre, B. *In vivo* regulation of the I κ B homologue cactus during the immune response of *Drosophila*. *J. Biol. Chem.* **273**, 10463–10469 (1998).
36. Fernandez, N. Q., Grosshans, J., Goltz, J. S. & Stein, D. Separable and redundant regulatory determinants in Cactus mediate its dorsal group dependent degradation. *Development* **128**, 2963–2974 (2001).
37. Drier, E. A., Huang, L. H. & Steward, R. Nuclear import of the *Drosophila* Rel protein Dorsal is regulated by phosphorylation. *Genes Dev.* **13**, 556–568 (1999).
38. Avila, A., Silverman, N., Diaz-Meco, M. T. & Moscat, J. The *Drosophila* atypical protein kinase C-ref(2)p complex constitutes a conserved module for signaling in the toll pathway. *Mol. Cell. Biol.* **22**, 8787–8795 (2002).
39. De Gregorio, E., Spellman, P. T., Rubin, G. M. & Lemaitre, B. Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. *Proc. Natl Acad. Sci. USA* **98**, 12590–12595 (2001).
40. Irving, P. *et al.* A genome-wide analysis of immune responses in *Drosophila*. *Proc. Natl Acad. Sci. USA* **98**, 15119–15124 (2001).
41. Tauszig, S., Jouanguy, E., Hoffmann, J. A. & Imler, J. L. Toll-related receptors and the control of antimicrobial peptide expression in *Drosophila*. *Proc. Natl Acad. Sci. USA* **97**, 10520–10525 (2000).
42. Kambis, Z., Hoffmann, J. A., Imler, J. L. & Capovilla, M. Tissue and stage-specific expression of the Tolls in *Drosophila* embryos. *Gene Expr. Patterns* **2**, 311–317 (2002).
43. Parker, J. S., Mizuguchi, K. & Gay, N. J. A family of proteins related to Spätzle, the toll receptor ligand, are encoded in the *Drosophila* genome. *Proteins* **45**, 71–80 (2001).
44. Keith, F. J. & Gay, N. J. The *Drosophila* membrane receptor Toll can function to promote cellular adhesion. *EMBO J.* **9**, 4299–4306 (1990).
45. Georgel, P. *et al.* *Drosophila* immune deficiency (IMD) is a death domain protein that activates antibacterial defense and can promote apoptosis. *Dev. Cell* **1**, 503–514 (2001).
46. Dushay, M. S., Asling, B. & Hultmark, D. Origins of immunity: Relish, a compound Rel-like gene in the antibacterial defense of *Drosophila*. *Proc. Natl Acad. Sci. USA* **93**, 10343–10347 (1996).
47. Hedengren, M. *et al.* Relish, a central factor in the control of humoral but not cellular immunity in *Drosophila*. *Mol. Cell* **4**, 827–837 (1999).
48. Stoven, S., Ando, I., Kadalayil, L., Engstrom, Y. & Hultmark, D. Activation of the *Drosophila* NF- κ B factor Relish by rapid endoproteolytic cleavage. *EMBO Rep.* **1**, 347–352 (2000).
49. Naïtza, S. *et al.* The *Drosophila* immune defense against Gram-negative infection requires the death protein dFADD. *Immunity* **17**, 575–581 (2002).
50. Leulier, F., Vidal, S., Saigo, K., Ueda, R. & Lemaitre, B. Inducible expression of double-stranded RNA reveals a role for dFADD in the regulation of the antibacterial response in *Drosophila* adults. *Curr. Biol.* **12**, 996–1000 (2002).
51. Vidal, S. *et al.* Mutations in the *Drosophila* dTAK1 gene reveal a conserved function for MAPKKs in the control of rel/NF- κ B-dependent innate immune responses. *Genes Dev.* **15**, 1900–1912 (2001).
52. Rutschmann, S. *et al.* Role of *Drosophila* IKK γ in a Toll-independent antibacterial immune response. *Nature Immunol.* **1**, 342–347 (2000).
53. Lu, Y., Wu, L. P. & Anderson, K. V. The antibacterial arm of the *Drosophila* innate immune response requires an I κ B kinase. *Genes Dev.* **15**, 104–110 (2001).
54. Silverman, N. *et al.* A *Drosophila* I κ B kinase complex required for Relish cleavage and antibacterial immunity. *Genes Dev.* **14**, 2461–2471 (2000).
55. Leulier, F., Rodriguez, A., Khush, R. S., Abrams, J. M. & Lemaitre, B. The *Drosophila* caspase Dredd is required to resist Gram-negative bacterial infection. *EMBO Rep.* **1**, 353–358 (2000).
56. Elrod-Erickson, M., Mishra, S. & Schneider, D. Interactions between the cellular and humoral immune responses in *Drosophila*. *Curr. Biol.* **10**, 781–784 (2000).
57. Hu, S. & Yang, X. dFADD, a novel death domain-containing adapter protein for the *Drosophila* caspase DREDD. *J. Biol. Chem.* **275**, 30761–30764 (2000).
58. Stoven, S. *et al.* Caspase-mediated processing of the *Drosophila* NF- κ B factor Relish. *Proc. Natl Acad. Sci. USA* **100**, 5991–5996 (2003).
59. Boutros, M., Agaisse, H. & Perrimon, N. Sequential activation of signaling pathways during innate immune responses in *Drosophila*. *Dev. Cell* **3**, 711–722 (2002).
60. De Gregorio, E., Spellman, P. T., Zou, P., Rubin, G. M. & Lemaitre, B. The Toll and Imd pathways are the major regulators of the immune response in *Drosophila*. *EMBO J.* **21**, 2568–2579 (2002).
61. Michel, T., Reichhart, J. M., Hoffmann, J. A. & Royet, J. *Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. *Nature* **414**, 756–759 (2001).
62. Gobert, V. *et al.* Toll activation during bacterial infection in *Drosophila* requires the concomitant function of two distinct blood-borne recognition proteins. *Science* (submitted).
63. Yoshida, H., Kinoshita, K. & Ashida, M. Purification of a peptidoglycan recognition protein from hemolymph of the silkworm, *Bombyx mori*. *J. Biol. Chem.* **271**, 13854–13860 (1996).
64. Lee, W. J., Lee, J. D., Kravchenko, V. V., Ulevitch, R. J. & Brey, P. T. Purification and molecular cloning of an inducible Gram-negative bacteria-binding protein from the silkworm, *Bombyx mori*. *Proc. Natl Acad. Sci. USA* **93**, 7888–7893 (1996).
65. Kang, D., Liu, G., Lundstrom, A., Gelius, E. & Steiner, H. A peptidoglycan recognition protein in innate immunity conserved from insects to humans. *Proc. Natl Acad. Sci. USA* **95**, 10078–10082 (1998).
66. Ochiai, M. & Ashida, M. A pattern recognition protein for peptidoglycan. Cloning the cDNA and the gene of the silkworm, *Bombyx mori*. *J. Biol. Chem.* **274**, 11854–11858 (1999).
67. Werner, T. *et al.* A family of peptidoglycan recognition proteins in the fruit fly *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **97**, 13772–13777 (2000).
68. Ochiai, M. & Ashida, M. A pattern-recognition protein for β -1,3-glucan. The binding domain and the cDNA cloning of β -1,3-glucan recognition protein from the silkworm, *Bombyx mori*. *J. Biol. Chem.* **275**, 4995–5002 (2000).
69. Ligoxygakis, P., Pelte, N., Hoffmann, J. A. & Reichhart, J. M. Activation of *Drosophila* Toll during fungal infection by a blood serine protease. *Science* **297**, 114–116 (2002).
70. Gottar, M. *et al.* The *Drosophila* immune response against Gram-negative bacteria is mediated by a peptidoglycan recognition protein. *Nature* **416**, 640–644 (2002).
71. Choe, K. M., Werner, T., Stoven, S., Hultmark, D. & Anderson, K. V. Requirement for a peptidoglycan recognition protein (PGRP) in Relish activation and antibacterial immune responses in *Drosophila*. *Science* **296**, 359–362 (2002).
72. Ramet, M., Manfrulli, P., Pearson, A., Mathey-Prevot, B. & Ezekowitz, R. A. Functional genomic analysis of phagocytosis and identification of a *Drosophila* receptor for *E. coli*. *Nature* **416**, 644–648 (2002).
73. Takehana, A. *et al.* Overexpression of a pattern-recognition receptor, peptidoglycan-recognition protein-LE, activates imd/relish-mediated antibacterial defense and the phenoloxidase cascade in *Drosophila* larvae. *Proc. Natl Acad. Sci. USA* **99**, 13705–13710 (2002).
74. Cheng, X., Zhang, X., Pflugrath, J. W. & Studier, F. W. The structure of bacteriophage T7 lysozyme, a zinc amidase and an inhibitor of T7 RNA polymerase. *Proc. Natl Acad. Sci. USA* **91**, 4034–4038 (1994).
75. Liepinsh, E., Genereux, C., Dehareng, D., Joris, B. & Otting, G. NMR structure of *Citrobacter freundii* AmpD, comparison with bacteriophage T7 lysozyme and homology with PGRP domains. *J. Mol. Biol.* **327**, 833–842 (2003).
76. Mellroth, P., Karlsson, J. & Steiner, H. A scavenger function for a *Drosophila* peptidoglycan recognition protein. *J. Biol. Chem.* **278**, 7059–7064 (2003).
77. Kim, M. S., Byun, M. & Oh, B. H. Crystal structure of peptidoglycan recognition protein LB from *Drosophila melanogaster*. *Nature Immunol.* **4**, 787–793 (2003).
78. Leulier, F. *et al.* The *Drosophila* immune system detects bacteria through specific peptidoglycan recognition. *Nature Immunol.* **4**, 478–484 (2003).
79. Liu, C., Xu, Z., Gupta, D. & Dziarski, R. Peptidoglycan recognition proteins: a novel family of four human innate immunity pattern recognition molecules. *J. Biol. Chem.* **276**, 34686–34694 (2001).
80. Dziarski, R., Platt, K. A., Gelius, E., Steiner, H. & Gupta, D. Defect in neutrophil killing and increased susceptibility to infection with non-pathogenic Gram-positive bacteria in peptidoglycan recognition protein-S (PGRP-S)-deficient mice. *Blood* **102**, 689–697 (2003).
81. Takeuchi, O. *et al.* Differential roles of TLR2 and TLR4 in recognition of Gram-negative and Gram-positive bacterial cell wall components. *Immunity* **11**, 443–451 (1999).
82. Gutierrez, O. *et al.* Induction of Nod2 in myelomonocytic and intestinal epithelial cells via nuclear factor- κ B activation. *J. Biol. Chem.* **277**, 41701–41705 (2002).
83. Girardin, S. E. *et al.* Nod1 detects a unique muropeptide from Gram-negative bacterial peptidoglycan. *Science* **300**, 1584–1587 (2003).
84. Becker, M. N., Diamond, G., Verghese, M. W. & Randell, S. H. CD14-dependent lipopolysaccharide-induced β -defensin-2 expression in human tracheobronchial epithelium. *J. Biol. Chem.* **275**, 29731–29736 (2000).
85. Liu, L., Roberts, A. A. & Ganz, T. By IL-1 signaling, monocyte-derived cells dramatically enhance the epidermal antimicrobial response to lipopolysaccharide. *J. Immunol.* **170**, 575–580 (2003).

Acknowledgements The author acknowledges invaluable exchanges with C. A. Janeway, R. A. Ezekowitz, J. Kafatos, T. Ganz and B. Beutler over many years, as well as the strongly motivating influence of H. G. Boman in the early period of these studies. I am particularly indebted to present and former associates of this group, namely P. Bulet, J. L. Dimarcq, D. Ferrandon, C. Hetru, D. Hoffmann, J. L. Imler, M. Lagueux, B. Lemaitre, E. Levashina, M. Meister, J. M. Reichhart and J. Royet.

Competing interests statement The author declares that he has no competing financial interests.

Correspondence and requests for materials should be addressed to J.A.H. (J.Hoffmann@ibmc.u-strasbg.fr).

Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside

Eva Pebay-Peyroula¹, Cécile Dahout-Gonzalez², Richard Kahn¹, Véronique Trézéguet³, Guy J.-M. Lauquin³ & Gérard Brandolin²

¹Institut de Biologie Structurale, UMR 5075 CEA-CNRS-Université Joseph Fourier, 41 rue Jules Horowitz, F-38027, Grenoble cedex 1, France

²Laboratoire de Biochimie et Biophysique des Systèmes Intégrés, UMR 5092 CEA-CNRS-Université Joseph Fourier, Département de Réponse et Dynamique Cellulaires, CEA-Grenoble, 17 Avenue des Martyrs, F-38054 Grenoble cedex 9, France

³Laboratoire de Physiologie Moléculaire et Cellulaire, UMR 5095 CNRS, Institut de Biochimie et Génétique Cellulaires, 1 rue Camille Saint-Saëns, F-33077 Bordeaux cedex, France

ATP, the principal energy currency of the cell, fuels most biosynthetic reactions in the cytoplasm by its hydrolysis into ADP and inorganic phosphate. Because resynthesis of ATP occurs in the mitochondrial matrix, ATP is exported into the cytoplasm while ADP is imported into the matrix. The exchange is accomplished by a single protein, the ADP/ATP carrier. Here we have solved the bovine carrier structure at a resolution of 2.2 Å by X-ray crystallography in complex with an inhibitor, carboxyatractyloside. Six α -helices form a compact transmembrane domain, which, at the surface towards the space between inner and outer mitochondrial membranes, reveals a deep depression. At its bottom, a hexapeptide carrying the signature of nucleotide carriers (RRRMMM) is located. Our structure, together with earlier biochemical results, suggests that transport substrates bind to the bottom of the cavity and that translocation results from a transient transition from a 'pit' to a 'channel' conformation.

The transport of various metabolites (nucleotides, phosphate, pyruvate, oxoglutarate, and so on) across the mitochondrial membranes is essential for eukaryotic metabolism. Specific transport through the inner mitochondrial membrane is achieved by nuclear encoded carriers which form a large transport family (the mitochondrial carrier family, MCF)¹. The exchange of ADP and ATP generates vital energy: each human being exchanges the equivalent of his/her own mass of ATP every day. The regeneration of ATP in mitochondria thus needs an efficient machinery able to import ADP and to export ATP. This transport is achieved by a membrane protein, the ADP/ATP carrier.

Severe diseases such as mitochondrial myopathies or ophthalmoplegia have been associated with dysfunctioning of the human ADP/ATP carrier^{2–6}. Translocation is very specifically inhibited by bongkrekic acids, secreted by the bacteria *Pseudomonas cocovenenans*, atractyloside (ATR) and carboxyatractyloside (CATR), of which the last is relevant here. Atractylosides are lethal poisons produced by *Atractylis gummifera*, a Mediterranean thistle known to the ancient Egyptians and described by the Greek physician Dioscorides (~40–90 AD) as a medicinal herb (*Chamaeleon Leukos*). More recently ATR and CATR were also identified in other plants^{7,8}.

The ADP/ATP carrier was discovered four decades ago^{9–11} and extensively studied^{12,13}. The 297 residues of the bovine carrier have been determined by protein sequencing¹⁴. It has been proposed that the carriers of various eukaryotic organisms share a high degree of sequence similarity as well as biochemical or functional properties. Human and bovine isoforms 1, localized in cardiac and skeletal muscles, share more than 90% amino acid sequence identity. Yeast and bovine carriers share about 50% identity. All ADP/ATP carriers exhibit a consensus sequence, RRRMMM, that is absent from other mitochondrial carriers. The two families of inhibitors provide tools for characterizing the nucleotide carrier. They were used in fluorescence measurements¹⁵ and in investigation of transport electrogenicity¹⁶. The results suggest that CATR binding occurs at a site similar to the ADP-binding site, thus preventing ADP/ATP transport.

With respect to the sequence, the carrier was predicted to consist of six transmembrane helices¹. The hydrophobicity of the second

transmembrane helix was found to be rather low¹⁷, and loop protrusions have been suggested^{18–20}. Both the amino and carboxy termini were shown to be oriented towards the intermembrane space^{21,22}. In functional terms, it was found that, as in most of the carriers, the stoichiometry of exchange of metabolites is one-to-one. The carrier appears to exist in two conformations which are involved in the translocation mechanism: one conformation is stabilized by atractylosides, the second by bongkrekic acid. It is now generally accepted that the functional carrier unit is a dimer. This was deduced from the oligomeric state of the CATR-inhibited carrier, which is a dimer with one CATR per dimer^{23,24}.

The structures at atomic resolution of several components of oxidative phosphorylation, such as ATP synthase and complexes II–IV of the electron transfer chain have been achieved. More recently, some structures of membrane proteins implicated in transport processes were reported (ion channelling, water or glycerol permeation, transport of small molecules and multidrug efflux) unveiling important mechanistic aspects (for a complete list of membrane protein structures, see http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html). To provide an insight into the nucleotide transport mechanism, an understanding of the structure of the carrier at high resolution is necessary. We now report this structure, determined by X-ray crystallography, to a resolution of 2.2 Å. Since ADP/ATP carrier is a paradigm for mitochondrial carriers, our results provide the first structural insight into the family of mitochondrial metabolite transporters.

Structure and overall architecture

The crystal unit cell contains one monomer per asymmetric unit and the crystal is packed as protein layers without indication of specific dimerization within the layers. Aminoacyl residues 2 to 293 of the molecule could be located in the model, and in addition, 82 molecules of water, one CATR and four lipids were seen. The transmembrane domain consists of six transmembrane α -helices, all of which are tilted relative to the orthogonal direction of the membrane and to each other (Fig. 1). The six helices form a barrel that defines a deep cone-shaped depression accessible from the

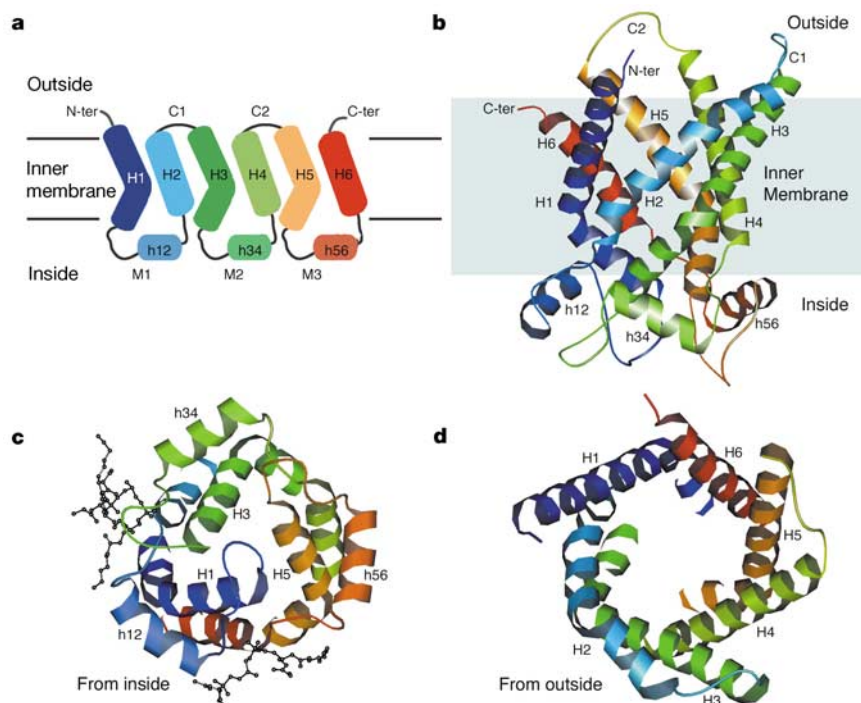


Figure 1 Architecture of the ADP/ATP carrier. **a**, A schematic diagram of the carrier secondary structure. Transmembrane helices, surface helices, intermembrane space loops and matrix loops are labelled H, h, C or M, respectively. Helices comprise the following residues: 4–37 (H1), 53–64 (h12), 73–99 (H2), 108–142 (H3), 156–167 (h34), 176–199 (H4), 209–238 (H5), 253–264 (h56), 273–290 (H6). Odd-numbered helices are kinked by the presence of prolines (P27, P132, P229). Inside and outside designate the matrix and the intermembrane space of mitochondria, respectively. **b**, A ribbon

diagram viewing the carrier from the side. The structure is coloured according to the sequence blue (N terminus) to red (C terminus). Membrane boundaries are drawn in agreement with the hydrophobic segments of the helices. **c**, View from the inside. Two cardiolipins are represented in black as ball and sticks. **d**, View from the outside. These and all subsequent ribbon diagrams were drawn with the programs MOLSCRIPT and BOBSCRIPT⁴⁹; surfaces with the program GRASP⁵⁰.

outside. The cavity has a maximal diameter of 20 Å and a depth of 30 Å (Fig. 2a). As for all mitochondrial carriers, the gene encoding ADP/ATP carrier presents a threefold duplication that results in a threefold repeat of about 100 amino acids within the protein¹⁷.

Figure 3a shows the alignment of the three repeats, which are related by less than 15% sequence identity. The fold of the three repeats is very similar (Fig. 3b), with root-mean-square deviations on the main chains of less than 2 Å. The connections within pairs of odd- and even-numbered helices contain short α -helical stretches (h12, h34 and h56). Situated parallel to the membrane surface, they appear to strengthen the closed conformation of the carrier on the matrix side (Fig. 1c). The surfaces that are oriented towards the membrane are hydrophobic and possibly interact with the membrane²⁵. Each of the odd-numbered helices exhibits a sharp kink, which is due to a proline residue (Fig. 1c and 3b). These prolines are located in the conserved sequence PX(D/E)XX(K/R) characteristic of mitochondrial carriers^{1,26}. Exposed on the outside, two short loops connect the three repeats to each other. The structure presents large hydrophilic surfaces, in particular in the interior of the conical pit, explaining the relatively weak hydrophobicity of the transmembrane helices, shown particularly for H2 (ref. 17). The electrostatic potential surface highlights strong positive patches at the bottom of the cavity (Fig. 4a) and weaker positive ones on the matrix side (Fig. 4b). The electron density map, $2F_{\text{obs}} - F_{\text{calc}}$, as well as the residual map, $F_{\text{obs}} - F_{\text{calc}}$, reveal cardiolipins in two locations on the matrix side (Fig. 1c), consistent with previous observations describing the presence of a few cardiolipins tightly bound to the carrier²⁷.

ADP/ATP transport blocked by trapping of CATR in the cavity

The distribution of charges and aromatic residues is asymmetric within the cavity, with a surface near the bottom that is very

hydrophilic. Towards the cavity, a cationic cluster consisting of five arginines (R79, R137, R234, R235, R279) and two lysines (K22, K32) points to the solvent. The positions of their side chains are stabilized by an intricate hydrogen-bond network, and by electrostatic interactions involving acidic or polar side chains (E29, D134, D231, Q36, N276 and so on) and water molecules. A second cationic region located near the entrance of the pit comprises residues from helix H2 (K91, K95), loop C1 (R104, K106) and helix H4 (R187, K198). The two basic patches are about 10 Å apart. In the cavity, a tyrosine motif is formed by Y186, Y190 and Y194 located on H4. Their aromatic rings are stacked, pointing with their

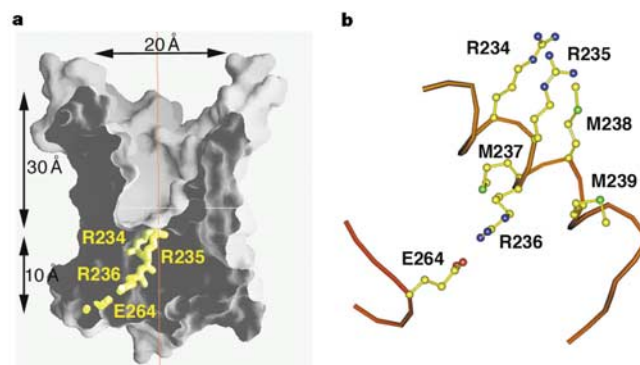


Figure 2 Section through the carrier. At the bottom of the cone-shaped cavity, the hexapeptide (RRMM signature) can be seen. **a**, The conical pit open to the outside and the RRR sequence spanning through the closed part of the carrier. **b**, A close-up view of the RRRMMM motif in the same orientation.

hydroxyl groups towards the centre of the cavity, thus forming a tyrosine ladder.

Initial experimental electron density maps showed extra density within the pit. After construction and refinement of the peptide chain, a CATR molecule (Fig. 5a–c) could be modelled unambiguously into these maps. CATR is located deeply in the cavity, not centred on the pseudo-threefold axis (Fig. 5b), and trapped through numerous interactions (Fig. 5c, d). The diterpene moiety is inserted towards the matrix side and stacks on the aromatic ring of Y186, while the glucose moiety carrying the two sulphate groups is exposed towards the outside. The tight binding of CATR at the bottom of the cavity is explained by the many hydrogen bonds that, in addition to water molecules, involve most of the basic residues previously mentioned. To determine the importance of the different chemical moieties of CATR, the inhibitory effect of various analogues of CATR have been studied²⁸.

In our structure, the isovaleric group is in van der Waals contact with hydrophobic side chains (L127, V130, I183). This is consistent with a affinity for ATR of the L142S mutant (equivalent to L127) of *Saccharomyces cerevisiae* ADP/ATP carrier (isoform 2) that is 15 times lower than for the wild type²⁹. The sulphate groups are within hydrogen-bond distances of several polar side chains (N87, K91, R187) as well as with the oxygen from the isovaleric residue for one sulphate and the oxygen of the hydroxyl linked to the sugar ring for the second sulphate. The two CATR carboxylic groups interact with the protein. The carboxylate that is common to CATR and ATR forms a salt bridge with R79. The additional carboxylate, characteristic of CATR, interacts with R279 via a water molecule (Fig. 5d), and therefore reinforces the affinity of CATR compared to ATR (the dissociation constant K_d is in the nanomolar range for CATR and is ten times higher for ATR). The primary alcohol group of the glucose moiety can be replaced by various chemical adducts without altering the inhibitory properties. Indeed, this group does not interact with the protein.

In summary, the structure shows that ADP/ATP carrier provides a very tight binding pocket for CATR, explaining the high efficiency of this lethal poison. It also shows why all the chemical groups (except the primary alcohol group of the ribose moiety) present in CATR are important for its toxicity.

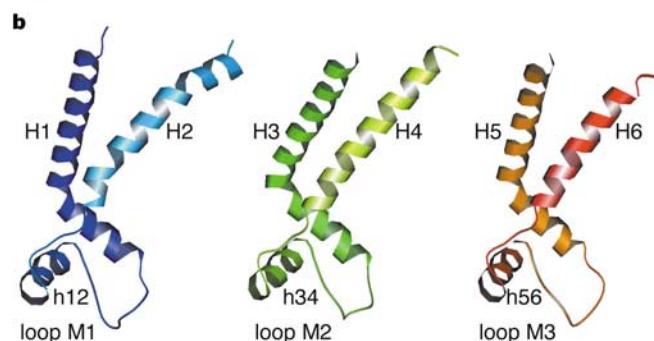
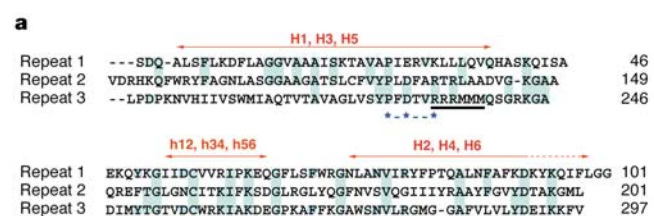


Figure 3 The threefold repeat of the ADP/ATP carrier. **a**, Sequence alignment of the three repeats. Conserved residues are highlighted in blue. Stars indicate the PX(D/E)XX(K/R) sequence. Helices are shown with arrows. The RRRMMM signature of nucleotide carriers is underlined in black. **b**, Structure comparison of the three repeats presented with similar orientations.

Discussion

The physiological role of the carrier is to export ATP from mitochondria in exchange for external ADP. Prior to transport, the carrier has to recognize the nucleotides and in contrast to most of nucleotide-binding proteins, it only binds adenine nucleotides that are not complexed with magnesium ions. Consensus nucleotide binding sequences such as Walker motifs are not present in the carrier. The ADP/ATP carrier is thought to function as a dimer that works as an antiport. Indeed, ADP/ATP transport has been demonstrated on kinetic grounds to proceed with a sequential mechanism accounting for the one-to-one stoichiometry of the exchange³⁰. In the absence of nucleotides, the carrier is thought to be multi-conformational. Each monomer within the dimer has two favourable conformations that can bind ADP or ATP from the outside or from the inside, respectively. Transport takes place once ADP binds to one monomer from the outside and ATP to the second monomer from the inside, simultaneously.

This antiport model resulting from various biochemical data raises several questions. What are the structural elements or amino acids involved in the nucleotide binding from each side? How flexible must the protein be to achieve the conformational changes associated with nucleotide binding and to allow the antiport process to take place? How are these changes triggered? We now discuss these questions in the light of our structure.

Structural basis for nucleotide binding

CATR binding is exclusive to nucleotide binding, raising the question of whether the binding sites are the same. Several genetic and biochemical approaches have aimed at identifying the residues that are important for nucleotide binding. In the *S. cerevisiae* carrier, the R96H mutation (named *op1*) drastically diminishes ADP binding affinity^{31,32}. Therefore, the bovine R79 residue (R96 in yeast) is most probably implicated in the binding of ADP. All the basic residues located at the bottom of the cavity in the bovine protein are highly conserved and therefore participate putatively in the binding of the negatively charged nucleotides. Indeed the role of many basic residues of the yeast carrier was investigated by mutagenesis. Arginines 96 (79), 204 (187), 252 (234), 253 (235), or 294 (279) as well as lysine 38 (22) (the corresponding residues of the bovine carrier are in parentheses) were shown to be essential for the transport activity³³, in contrast to K179 (162) and K182 (165) that can be mutated concomitantly into isoleucines without affecting the function³⁴.

In our structure, the residues shown to be important for the activity are interacting with CATR and are located within the putative nucleotide-binding site. By analogy to the van der Waals distance between the diterpene moiety of CATR and Y186, the adenine ring could interact with Y186, a highly conserved residue within the ADP/ATP carrier family. Similar types of interactions

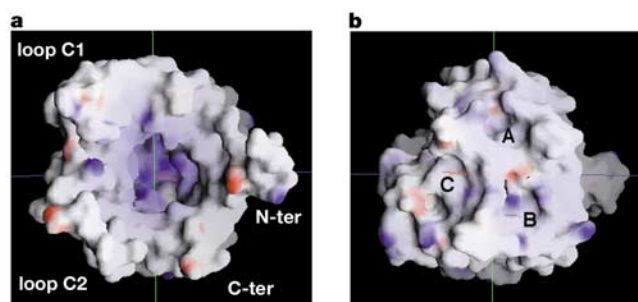


Figure 4 Electrostatic potential surface. Positive and negative surfaces are shown in blue and red, respectively. **a**, View from the outside. **b**, View from the inside showing three small cavities (A, B, C). Two cationic patches are located in the regions labelled B (R139/K162/K165) and C (R236/R243/R258/K259).

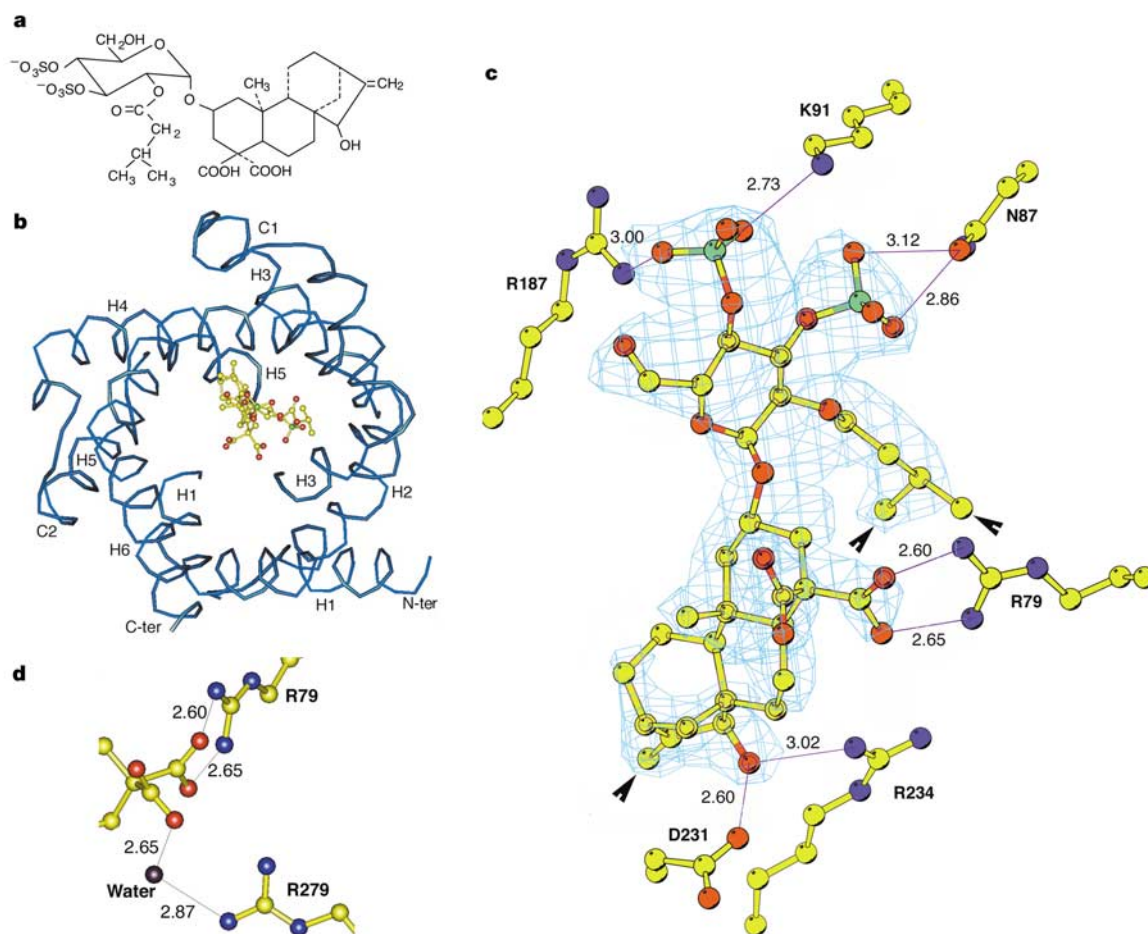


Figure 5 The binding of CATR. **a**, Chemical formula of CATR. **b**, CATR within the binding pocket viewed from the outside. **c**, Electron density map at 2.2 Å resolution ($2F_{\text{obs}} - F_{\text{calc}}$, contoured at 1σ obtained before CATR modelling and including only protein atoms) around CATR. Residues directly within hydrogen-bond distances to CATR

are labelled. Black arrowheads indicate van der Waals contacts between CATR and the protein. **d**, A close-up of the two CATR carboxylates interacting with arginines. Interatomic distances are given in Å.

were observed in the structures of several ATP-binding proteins³⁵. Altogether, these facts support the idea that ADP and CATR binding sites overlap. In addition, structural features hint how nucleotides could be guided towards the binding site. The patch of basic residues described at the entrance of the surmised channel could participate in attracting nucleotides towards their binding site. A stacking interaction with tyrosine rings could guide the ADP translocation along the tyrosine ladder (Y186, Y190, Y194). In yeast, we have shown that double mutations of Y203 and Y207 (equivalent to Y186 and Y190 in the bovine carrier) prevented or slightly reduced cellular growth on a non-fermentable carbon source when mutated into alanines or phenylalanines, respectively (data not shown).

RRRMMM may trigger conformational change

The conformation of the carrier required for ATP binding from the matrix is probably different from the ADP-binding conformation and therefore, assessment of the ATP binding site is more speculative. Individual mutations of arginines in the RRRMMM motif abolish the transport activity. The three arginines span over the apparent barrier, depicted in Fig. 2a, that the nucleotide has to cross during translocation through the membrane. R234 and R235 are implicated in CATR binding and therefore possibly in ADP binding. R236, which points in the opposite direction (Fig. 2b), forms a salt bridge with E264 (also conserved among the nucleotide carriers), contributing to the interaction of M3 with the core of the protein.

Matrix loop M3 contains a sequence of eight amino acids including charged amino acids E264, K267 and K271. It was reported that Asp or Glu carboxylates are possible partners in hydrogen bonds with adenine or ribose moieties, while lysine N ζ nitrogens interact frequently with phosphate oxygens^{35,36}. The salt bridge involving E264 with R236 hides both side chains from the surface, impairing potential nucleotide binding. On the basis of the structure, we propose RRR to be a two-way switch that dictates nucleotide binding stoichiometry. ATP coming from the matrix might destabilize the salt bridge, inducing a local rearrangement of the RRR structure due to the reorientation of E264 and thus impeding the accessibility of R234 and R235 to external nucleotides. Conversely, binding of ADP from the outside would mask R236 and prevent nucleotide binding from the matrix. In the structure MMM occupies a bulky volume (Fig. 2b) that could control the access, acting as a plug during the transport mechanism. In addition, a possible reorientation of E264 that favoured ATP interaction with the glutamate via its adenine or ribose moiety could also favour ATP interaction with K267 or K271 via the phosphate groups.

Hinges and channel opening

Figure 6 depicts the conformation of odd-numbered helices and shows how the prolines of the PX(D/E)XX(K/R) sequence impose a kink on them. Three pairs of charged residues strengthen a closed conformation by multiple interactions. We propose that during the transport process prolines act as hinges that would straighten odd-

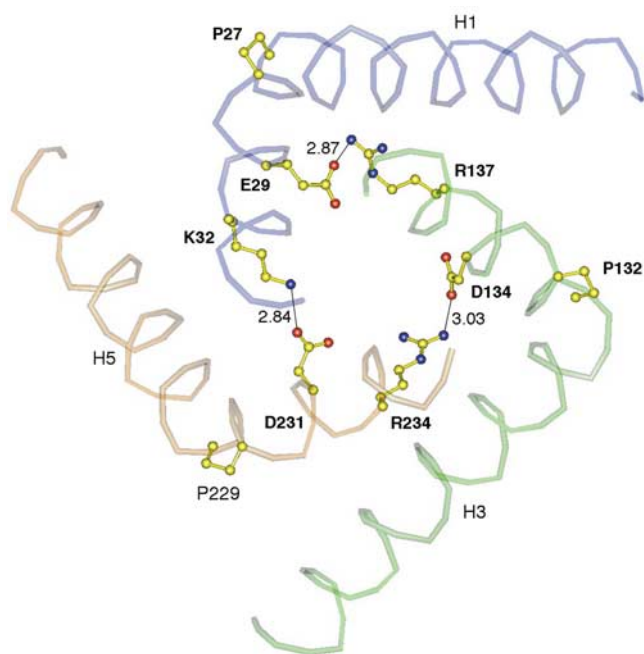


Figure 6 The closed conformation of the carrier viewed from the matrix. Only odd-numbered helices are represented. Residues involved in the PX(D/E)XX(K/R) sequence are labelled. Black lines connect atoms within hydrogen-bonding distances (in Å).

numbered helices, thus reshaping the protein and pulling open a channel. This would also imply a reorganization of charged pairs (Fig. 6) as previously suggested²⁶. Opening and closing channels through helix bending was also proposed for other proteins^{37,38}. Once channels are formed, the vectoriality of the transport within the ADP/ATP carrier is probably guided by the electrostatic environment.

To prevent binding of two nucleotides from the same side on each monomer within a dimer, we postulate that dimeric structural constraints require the conformations of both monomers to be out-of-phase. Therefore, binding of one nucleotide from either side of the carrier favours binding of a second one from the opposite side, consistent with previously reported biochemical data³⁹. Simultaneous binding of both nucleotides could impose lateral forces on odd-numbered helices, modifying their bending as described above. Many features of the structure disrupt the threefold symmetry of the carrier monomer structure. In particular, M1 is connected to the core of the protein through a number of hydrogen-bonding and electrostatic interactions that is significantly larger than for M2 or M3. These observations raise the question of whether the three matrix loops follow a symmetric movement during the transport.

Conclusion

The structure of the ADP/ATP carrier complexed with CATR provides an insight into one conformation of the protein. It provides the molecular basis of the high-affinity binding of CATR and excludes CATR binding between two monomers, as previously proposed. Even though the structure that we describe here represents a monomer, key structural determinants that are important in nucleotide transport can be learned from it. RRRMMM, the ADP/ATP carrier signature, spans over the thinnest part of the protein in a strategic location for the transport. Attraction of ADP towards the matrix against the natural electrostatic potential is highlighted by the distribution of positive charge patches within the protein cavity. The carrier has a closed conformation on the matrix side induced by kinks in odd-numbered helices due to the presence of prolines. We hypothesize that the transport itself could be

triggered by cooperative binding of nucleotides and achieved through channels that open as the odd-numbered transmembrane helices un-tilt with respect to the membrane, with the prolines acting as hinges. To sustain this hypothesis the different conformations have to be explored by various methods including two-dimensional electron microscopy⁴⁰ or X-ray crystallography. Because the PX(D/E)XX(K/R) sequence is characteristic of all mitochondrial carriers, the opening mechanism could be extended to other transporters, the transition from open to closed states being specifically triggered by the binding of transported metabolites.

It is also probable that ADP/ATP transport in mitochondria cannot be described fully at the molecular level only: it may well result from a combination of stochastic events and parameters of the local environment, including membrane potential and nucleotide concentrations on both sides of the inner membrane. Although our model sheds light on some features of the mechanism, the nucleotide selectivity (adenine versus guanine), the transport itself once nucleotides are bound, and kinetic aspects need additional experimental investigations to be deciphered. □

Methods

Detailed information on methodological aspects is provided as Supplementary Information.

Purification and crystallization

The ADP/ATP carrier (relative molecular mass, $M_w = 33,000$ for the monomer) was isolated from beef heart mitochondria as a CATR-carrier complex in the presence of the detergent 3-laurylamido-N,N'-dimethylpropylaminoxide (LAPAO) according to the procedure of ref. 24 with minor modifications. The purified carrier fraction was treated with Bio-Beads (BioRad) to remove excess detergent⁴¹ before concentration. Crystals were grown by hanging-drop vapour diffusion at 20 °C by mixing equal volumes of protein and reservoir. Two different orthorhombic crystal forms were obtained in the presence of 30% Jeffamine-M600 (O-(2-aminopropyl)-O'-(2-methoxyethyl)polypropylene glycol 500) as a precipitant. The first crystals are primitive ($P2_12_12_1$, $a = 85.4$ Å, $b = 83.5$ Å, $c = 49.9$ Å) and the second ones centred ($C222_1$, $a = 79.9$ Å, $b = 109.2$ Å, $c = 89.3$ Å). The primitive crystal form used for the refinement contains monomers probably due to selective crystallization of the CATR-carrier monomeric complex. Excessive delipidation could account for dimer dissociation. Indeed, it was reported that lipids contribute strongly to oligomerization as shown in bacteriorhodopsin⁴² or bacterial cytochrome *c* oxidase⁴³.

Structure determination

Numerous data sets were collected at the European Synchrotron Radiation Facility, Grenoble (ESRF), on beam lines BM30A, ID14-1, ID14-2 and ID29, at 100 K under nitrogen stream. Data sets used for phasing and refinement were collected on beamlines ID29 and BM30A. All data sets were processed with DENZO, SCALEPACK⁴⁴ and CCP4⁴⁵. Most of the primitive crystals presented a defect and were not usable, with the exception of one good native data set to 2.2 Å that was used for the model refinement. The heavy-atom derivatives necessary for the phase determination were screened with the centred form despite anisotropic diffraction. Three mercury derivatives were collected and used in phase determination with phasing powers of 1.80, 1.58 and 1.70, respectively. Heavy-atom site positions deduced manually from Patterson maps were refined and phases calculated using MLPHARE⁴⁵. Initial phases were improved by solvent flattening using DM⁴⁵, increasing the figure of merit from 0.55 to 0.69. The model was built using O⁴⁶ and refinement was carried out using CNS⁴⁷. The electron density map clearly showed six helices. A first model containing the helices and part of the loops was constructed in the centred crystal form and transposed to the primitive form by molecular replacement⁴⁸. The model construction was improved using both the experimental electron density transposed to the primitive crystal form according to the rotation matrix and translation vector obtained by molecular replacement and the $2F_{\text{obs}} - F_{\text{calc}}$ maps. The model was refined by iterative cycles of manual corrections and energy minimization or simulated annealing followed by individual *B*-factor refinements, to a final R_{cryst} of 22.6% and R_{free} of 26.8% to 2.2 Å resolution.

Received 27 June; accepted 15 September 2003; doi:10.1038/nature02056.

- Walker, J. E. & Runswick, M. J. The mitochondrial transport protein superfamily. *J. Bioenerg. Biomembr.* **25**, 435–446 (1993).
- Torroni, A., Stepien, G., Hodge, J. A. & Wallace, D. C. Neoplastic transformation is associated with coordinate induction of nuclear and cytoplasmic oxidative phosphorylation genes. *J. Biol. Chem.* **265**, 20589–20593 (1990).
- Heddi, A., Lestienne, P., Wallace, D. C. & Stepien, G. Mitochondrial DNA expression in mitochondrial myopathies and coordinated expression of nuclear genes involved in ATP production. *J. Biol. Chem.* **268**, 12156–12163 (1993).
- Kaukonen, J. *et al.* Role of adenine nucleotide translocator 1 in mtDNA maintenance. *Science* **289**, 782–785 (2000).
- Fiore, C., Arlot-Guilligay, D., Trézéguet, V., Lauquin, G. J. & Brandolin, G. Fluorometric detection of ADP/ATP carrier deficiency in human muscle. *Clin. Chim. Acta* **311**, 125–135 (2001).
- Napoli, L. *et al.* A novel missense adenine nucleotide translocator-1 gene mutation in a Greek adPEO family. *Neurology* **57**, 2295–2298 (2001).

7. Craig, J. C., Mole, M., Billets, S. & El-Feraly, F. Isolation and identification of hypoglycemic agent carboxyatractylate from *Xanthium strumarium*. *Phytochemistry* **15**, 1178 (1976).
8. Candy, H. A., Pegel, K. H., Brookes, B. & Rodwell, M. The occurrence of atractylolide in *Callilepis laureola*. *Phytochemistry* **16**, 1308–1309 (1977).
9. Bruni, A., Luciani, S. & Contessa, A. R. Inhibition by atractylolide of the binding of adenine nucleotides of rat liver mitochondria. *Nature* **201**, 1219–1220 (1964).
10. Duée, E. D. & Vignais, P. V. Exchange between extra- and intramitochondrial adenine nucleotides. *Biochim. Biophys. Acta* **107**, 184–188 (1965).
11. Pfaff, E., Klingenberg, M. & Heldt, H. W. Unspecific permeation and specific exchange of adenine nucleotides in liver mitochondria. *Biochim. Biophys. Acta* **104**, 312–315 (1965).
12. Klingenberg, M. Molecular aspects of the adenine nucleotide carrier from mitochondria. *Arch. Biochem. Biophys.* **270**, 1–14 (1989).
13. Fiore, C. *et al.* The mitochondrial ADP/ATP carrier: structural, physiological and pathological aspects. *Biochimie* **80**, 137–150 (1998).
14. Aquila, H., Misra, D., Eulitz, M. & Klingenberg, M. Complete amino acid sequence of the ADP/ATP carrier from beef heart mitochondria. *Hoppe-Seyler's J. Physiol. Chem.* **363**, 345–349 (1982).
15. Brandolin, G., Dupont, Y. & Vignais, P. V. Substrate-induced modifications of the intrinsic fluorescence of the isolated adenine nucleotide carrier protein: demonstration of distinct conformational states. *Biochemistry* **24**, 1991–1997 (1985).
16. Gropp, T. *et al.* Kinetics of electrogenic transport by the ADP/ATP carrier. *Biophys. J.* **77**, 714–726 (1999).
17. Walker, J. E. The mitochondrial transporter family. *Curr. Opin. Struct. Biol.* **2**, 519–526 (1992).
18. Bogner, W., Aquila, H. & Klingenberg, M. The transmembrane arrangement of the ADP/ATP carrier as elucidated by the lysine reagent pyridoxal 5-phosphate. *Eur. J. Biochem.* **161**, 611–620 (1986).
19. Majima, E., Koike, H., Hong, Y. M., Shinohara, Y. & Terada, H. Characterization of cysteine residues of mitochondrial ADP/ATP carrier with the SH-reagents eosin 5-maleimide and N-ethylmaleimide. *J. Biol. Chem.* **268**, 22181–22187 (1993).
20. Dianoux, A. C. *et al.* Two distinct regions of the yeast mitochondrial ADP/ATP carrier are photolabeled by a new ADP analogue: 2-azido-3'-O-naphthoyl-[β - 32 P]ADP. Identification of the binding segments by mass spectrometry. *Biochemistry* **39**, 11477–11487 (2000).
21. Brandolin, G., Boulay, F., Dalbon, P. & Vignais, P. V. Orientation of the N-terminal region of the membrane-bound ADP/ATP carrier protein explored by antipeptide antibodies and an arginine-specific endoprotease. Evidence that the accessibility of the N-terminal residues depends on the conformational state of the carrier. *Biochemistry* **28**, 1093–1100 (1989).
22. Trézéguet, V. *et al.* A covalent tandem dimer of the mitochondrial ADP/ATP carrier is functional *in vivo*. *Biochim. Biophys. Acta* **1457**, 81–93 (2000).
23. Hackenberg, H. & Klingenberg, M. Molecular weight and hydrodynamic parameters of the adenosine 5'-diphosphate/adenosine 5'-triphosphate carrier in TritonX-100. *Biochemistry* **19**, 548–555 (1980).
24. Block, M. R., Zaccari, G., Lauquin, G. J. & Vignais, P. V. Small angle neutron scattering of the mitochondrial ADP/ATP carrier protein in detergent. *Biochem. Biophys. Res. Commun.* **109**, 471–477 (1982).
25. Pannels, V., Schussler, U., Costagliola, S. & Sinning, I. Choline head groups stabilize the matrix loop regions of the ATP/ADP carrier ScaAC2. *Biochem. Biophys. Res. Commun.* **300**, 65–74 (2003).
26. Nelson, D. R., Felix, C. M. & Swanson, J. M. Highly conserved charge-pair networks in the mitochondrial carrier family. *J. Mol. Biol.* **277**, 285–308 (1998).
27. Beyer, K. & Klingenberg, M. ADP/ATP carrier protein from beef heart mitochondria has high amounts of tightly bound cardiolipin, as revealed by 31 P nuclear magnetic resonance. *Biochemistry* **24**, 3821–3826 (1985).
28. Vignais, P. V. Molecular and physiological aspects of adenine nucleotide transport in mitochondria. *Biochim. Biophys. Acta* **456**, 1–38 (1976).
29. Zeman, I. *et al.* Four mutations in transmembrane domains of the mitochondrial ADP/ATP carrier increase resistance to bongkreic acid. *J. Bioenerg. Biomembr.* **35**, 243–256 (2003).
30. Duyckaerts, C., Sluse-Goffart, C. M., Fux, J. P., Sluse, F. E. & Liebecq, C. Kinetic mechanism of the exchanges catalysed by the adenine-nucleotide carrier. *Eur. J. Biochem.* **106**, 1–6 (1980).
31. Kovac, L., Lachowicz, T. M. & Slonimski, P. P. Biochemical genetics of oxidative phosphorylation. *Science* **158**, 1564–1567 (1967).
32. Beck, J. C., Mattoon, J. R., Hawthorne, D. C. & Sherman, F. Genetic modification of energy-conserving systems in yeast mitochondria. *Proc. Natl Acad. Sci. USA* **60**, 186–193 (1968).
33. Nelson, D. R., Lawson, J. E., Klingenberg, M. & Douglas, M. G. Site-directed mutagenesis of the yeast mitochondrial ADP/ATP translocator. Six arginines and one lysine are essential. *J. Mol. Biol.* **230**, 1159–1170 (1993).
34. Müller, V., Heidkamper, D., Nelson, D. R. & Klingenberg, M. Mutagenesis of some positive and negative residues occurring in repeat triad residues in the ADP/ATP carrier from yeast. *Biochemistry* **36**, 16008–16018 (1997).
35. Denessiouk, K. A. & Johnson, M. S. When fold is not important: a common structural framework for adenine and AMP binding in 12 unrelated protein families. *Proteins* **38**, 310–326 (2000).
36. Moodie, S. L., Mitchell, J. B. & Thornton, J. M. Protein recognition of adenylate: an example of a fuzzy recognition template. *J. Mol. Biol.* **263**, 486–500 (1996).
37. Jiang, Y. *et al.* Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **417**, 515–522 (2002).
38. Abramson, J. *et al.* Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* **301**, 610–615 (2003).
39. Block, M. R. & Vignais, P. V. Substrate-site interactions in the membrane-bound adenine-nucleotide carrier as disclosed by ADP and ATP analogs. *Biochim. Biophys. Acta* **767**, 369–376 (1984).
40. Kunji, E. R. & Harding, M. Projection structure of the atractylolide-inhibited mitochondrial ADP/ATP carrier of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **278**, 36985–36988 (2003).
41. Holloway, P. W. A simple procedure for removal of Triton X-100 from protein samples. *Anal. Biochem.* **53**, 304–308 (1973).
42. Belrhali, H. *et al.* Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 Å resolution. *Struct. Fold. Des.* **7**, 909–917 (1999).
43. Harrenga, A. & Michel, H. The cytochrome c oxidase from *Paracoccus denitrificans* does not change the metal center ligation upon reduction. *J. Biol. Chem.* **274**, 33296–33299 (1999).
44. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
45. The CCP4 suite: programs for X-ray crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
46. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
47. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
48. Navaza, J. Implementation of molecular replacement in AMoRe. *Acta Crystallogr. D* **57**, 1367–1372 (2001).
49. Esnouf, R. M. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D* **55**, 938–940 (1999).
50. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the staff at the European Synchrotron Radiation Facilities and of the French beamline BM30A (ESRF) for synchrotron support. We also thank J. P. Rosenbusch and R. Douce for reading the manuscript, and for numerous suggestions and discussions, and P. V. Vignais for helpful comments. This work was supported by the programmes PCV (CNRS) and Emergence (Région Rhône-Alpes) and by the Région Aquitaine.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.P.P. (pebay@ibs.fr) or G.B. (gbrandolin@cea.fr). Coordinates have been deposited in the Protein Data Bank under accession code 1okc.

Voyager 1 exited the solar wind at a distance of ~ 85 AU from the Sun

S. M. Krimigis¹, R. B. Decker¹, M. E. Hill², T. P. Armstrong³, G. Gloeckler², D. C. Hamilton², L. J. Lanzerotti^{4,5} & E. C. Roelof¹

¹Applied Physics Laboratory, The Johns Hopkins University, Laurel, Maryland 20723, USA

²Department of Physics, University of Maryland, College Park, Maryland 20742, USA

³Fundamental Technologies, 2411 Ponderosa, Suite A, Lawrence, Kansas 66046, USA

⁴Bell Laboratories, 600 Mountain Avenue, Building 1E-439, Murray Hill, New Jersey 07974, USA

⁵Center for Solar Terrestrial Research, New Jersey Institute of Technology, Newark, New Jersey 07102, USA

The outer limit of the Solar System is often considered to be at the distance from the Sun where the solar wind changes from supersonic to subsonic flow¹. Theory predicts that a termination shock marks this boundary, with locations ranging² from a few to over 100 AU (1 AU $\approx 1.5 \times 10^8$ km, the distance from Earth to the Sun). 'Pick-up ions' that originate^{3,4} as interstellar neutral atoms should be accelerated to tens of MeV at the termination shock, generating anomalous cosmic rays⁵⁻⁷. Here we report a large increase in the intensity of energetic particles in the outer heliosphere, as measured by an instrument on the Voyager 1 spacecraft. We argue that the spacecraft exited the supersonic solar wind and passed into the subsonic region (possibly beyond the termination shock) on about 1 August 2002 at a distance of ~ 85 AU (heliolatitude $\sim 34^\circ$ N), then re-entered the supersonic solar wind about 200 days later at ~ 87 AU from the Sun. We show that the composition of the ions accelerated at the putative termination shock is that of anomalous cosmic rays and of interstellar pick-up ions.

The low-energy charged particle (LECP) instrument⁸ on Voyagers 1 and 2 consists of a collection of solid-state detectors designed to perform measurements of ions (~ 30 keV to tens of MeV) and electrons (~ 28 keV to ~ 10 MeV), including elemental composition over a range of energies (≥ 0.3 MeV per nucleon) and species (H, He, C, N, O, Ne, and so on). The double-ended detector telescope system is mounted on a stepping platform that rotates through 360° parallel to the R-T plane of the [R,T,N] coordinate system in eight sectors, with one sector shielded so as to provide background measurements. Each step occurs every 192 s.

In late 2000 (Fig. 1), a sustained increase of low-energy protons was measured at Voyager 1 with less intense but more structured activity measured at Voyager 2 (time-shifted for convection to the distance of Voyager 1 from the Sun ('radius') using the observed solar wind velocity at Voyager 2). Possible precursor enhancements in the 0.6 MeV protons occurred in late 2000 when Voyager 1 was located at ~ 80 AU. Only two Voyager 1 increases (on decimal years 2001.2 and 2000.0) before the 2002.5 event are correlated with Voyager 2, while other Voyager 1 increases show higher intensities for this period, contrary to the conclusion of ref. 9. In mid-2002, however, there was a large, 100-fold increase at Voyager 1 with no obvious counterpart at Voyager 2. The Voyager 1 increase is the largest since the 1991 solar-flare-associated event when the spacecraft was located approximately half its present distance from the Sun¹⁰. From 2002.5 to 2003.1 little correlation is evident in either peak intensities or duration between Voyager 1 and the Voyager 2 time-shifted profiles. The possibility remains, however, that if significant solar/interplanetary activity occurred predominantly in northern heliographic latitudes (where Voyager 1 is located), it might have propagated asymmetrically and reached Voyager 1

without being first observed by Voyager 2 at southern heliolatitudes. Observations from instruments on the Advanced Composition Explorer (ACE) (at 1 AU) in the ecliptic plane and Ulysses (at ~ 4 AU, ~ 24 – 45° N) argue against such a possibility. Measured intensities between 1 and 4 AU are found to decrease in helioradius as $\sim R^{-2}$ while those from Ulysses to Voyager 1 decline as $\sim R^{-1}$, that is, the Voyager 1 intensities could not have been convected from the inner heliosphere without additional acceleration¹¹. Finally, a Forbush decrease in galactic cosmic rays (GCR) typically associated with interplanetary events such as the one in 1991 (ref. 10) is not only absent in the last half of 2002, but replaced by a $\sim 20\%$ increase (Fig. 1c). Note the coherence at Voyager 1 between the GCRs and the lower-energy particles in both the overall increase as well as the relative maxima.

The earliest possible onset in mid-2002 is on day 162 (11 June 2002) when Voyager 1 was located at ~ 84.7 AU. Judging from typical reverse interplanetary shocks observed on Ulysses¹², however, these increases were probably precursors, whereas the exit from the solar wind most probably occurred near day 213 ± 5 days (1 August 2002). Importantly, spatial/temporal structures as short as 1–3 days exist, a highly unusual characteristic for either co-rotating inter-

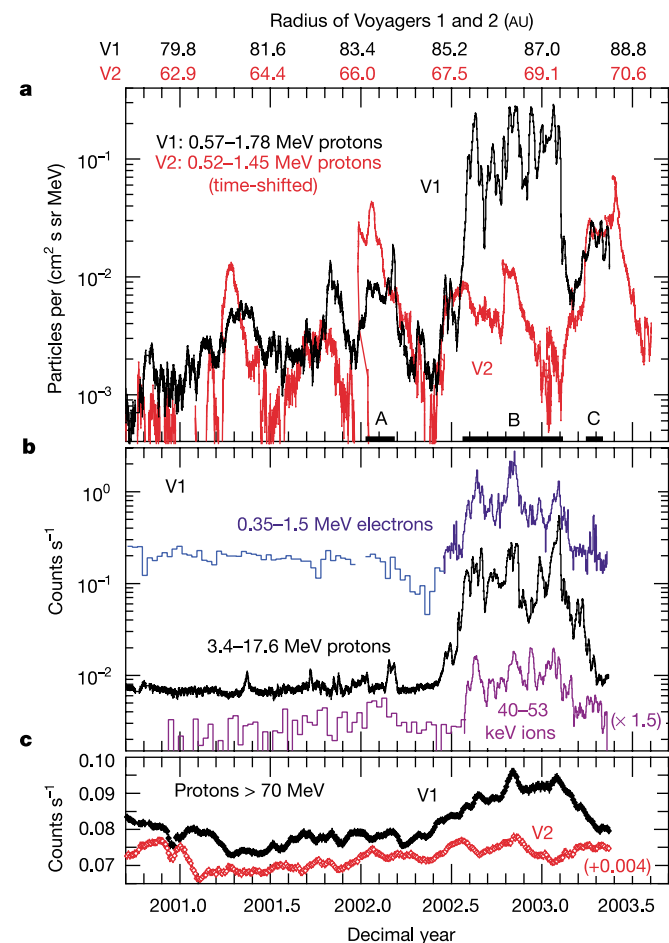


Figure 1 Intensity profiles at Voyager 1, Voyager 2 since late 2001. **a**, Profiles at Voyager 1 and Voyager 2 (time-shifted) using solar wind velocities measured at Voyager 2 (J. Richardson, personal communication). The Voyager 1 increase on 2002.5 is unique with no counterpart at Voyager 2. The three horizontal bars (A, B, C) denote periods where angular distributions were used to infer the solar wind velocity (Fig. 3). **b**, More-detailed profiles of daily averages for indicated energies and species. **c**, Intensity of >70 MeV GCRs at Voyager 1, together with equivalent channel (non-time-shifted) at Voyager 2.

action regions or global merged interaction regions¹³. Such short-term structures (Fig. 1a, b) are typically associated with shocks and are most probably either connection/disconnection of the interplanetary magnetic field with the shock fronts¹⁴ or possibly filamentary structures within this region of the heliosphere. The acceleration of high-energy electrons over such a long period in the outer heliosphere (as evident by the intensity increases in Fig. 1b) is only consistent with previous observations at perpendicular shocks¹⁵.

Anisotropy measurements for ~ 0.6 MeV protons are shown in Fig. 2. Protons arrive at the spacecraft predominantly from the azimuthal direction (sectors 7 and 6), between which the average Parker interplanetary magnetic field is expected to be located, and stream 'outward', implying a source inside the location of Voyager 1. The hemisphere in the opposite direction (sectors 1–4) displays a nearly isotropic distribution. This observation would be surprising if Voyager 1 was still inside the termination shock, since one would then expect a significantly higher intensity in sector 1 than sector 5 owing to convection from the solar direction by the solar wind. That prevalence of sector 1 rates has been observed by LECP for the past 25 years (ref. 16).

Figure 3 shows model fits (see Methods) and observed sector rates for the period of interest plus averaged rates during periods from before and after the mid-2000 event. Figure 3a shows fits to ~ 1 MeV protons before, during and after the 2002.5 increase, with solar wind velocities of ~ 300 , ~ 0 and ~ 230 km s⁻¹, respectively. Figure 3b displays similar fits in five consecutive energy intervals. The 0 km s⁻¹ model is in good agreement with the observations and the fitting error (Fig. 3c) suggests an upper limit of <50 km s⁻¹. It is also remarkable that the inferred direction of the interplanetary magnetic field is $\sim 22^\circ$ inward from the T-direction where the Parker spiral predicts a tangential field direction to within a degree at this distance in the solar wind. We conclude from this analysis that the plasma flow changed from supersonic ($\sim 300 \pm 30$ km s⁻¹) in early 2002 to subsonic (<50 km s⁻¹) for the period of high intensities and back to supersonic ($\sim 230 \pm 25$ km s⁻¹) following re-entry into the upstream solar wind, validating the premise that Voyager 1 had

crossed a boundary for the first time in the July–August timeframe of 2002 and again in early 2003.

The Voyager 1 energy spectra (Fig. 4) demonstrate that the low-energy enhancement is relatively poor in carbon ($C/O \approx 0.02$), as is the case for pick-up ions and anomalous cosmic rays (ACRs). The spectral shape is not the monotonically decreasing form for an unmodulated spectrum that was expected beyond the termination shock⁶. Instead, there appear to be two components—a power-law component at lower energies and a peaked spectrum at higher energies, observable at ~ 7 MeV per nucleon for O and indicated at the highest He energies shown. In this interpretation, the entire proton spectrum shown is locally accelerated at the termination shock, as are the He and O spectra below 10 MeV per nucleon and 2 MeV per nucleon, respectively. Above those energies the local He and O components are masked by an ACR component accelerated elsewhere. The spectral peaks thus result in the usual way, from energy-dependent transport from the remote acceleration sites to Voyager 1¹⁷.

The salient features of the observations are as follows: (1) A gradual increase in ~ 1 -MeV proton intensity in late 2000 culminated in a 100-fold step increase in mid-2002 lasting for over six months and ending within a few hours in early 2003. (2) The step increase was seen in energies ranging from ~ 40 keV to >70 MeV for protons and >0.35 MeV for electrons. (3) Large intensity fluctuations lasting from <1 to several days were present coherently in all energies and species throughout the period. (4) The composition of H, He, O and C are consistent with an ACR or pick-up ion source. (5) Outward-streaming anisotropies show that the particle source lies inside the Voyager 1 location. (6) Generalized Compton–Getting fits to the angular distributions show that the convection velocity changed from supersonic to subsonic and back to supersonic before, during, and after the increase, respectively.

The LECP data presented above demonstrate that Voyager 1 in July–August 2002 entered a new region in the outer heliosphere at a distance of ~ 85 AU, remained in it for about six months and re-entered the upstream solar wind at 87.4 AU. This conclusion is supported by the large, prolonged ion- and electron-intensity

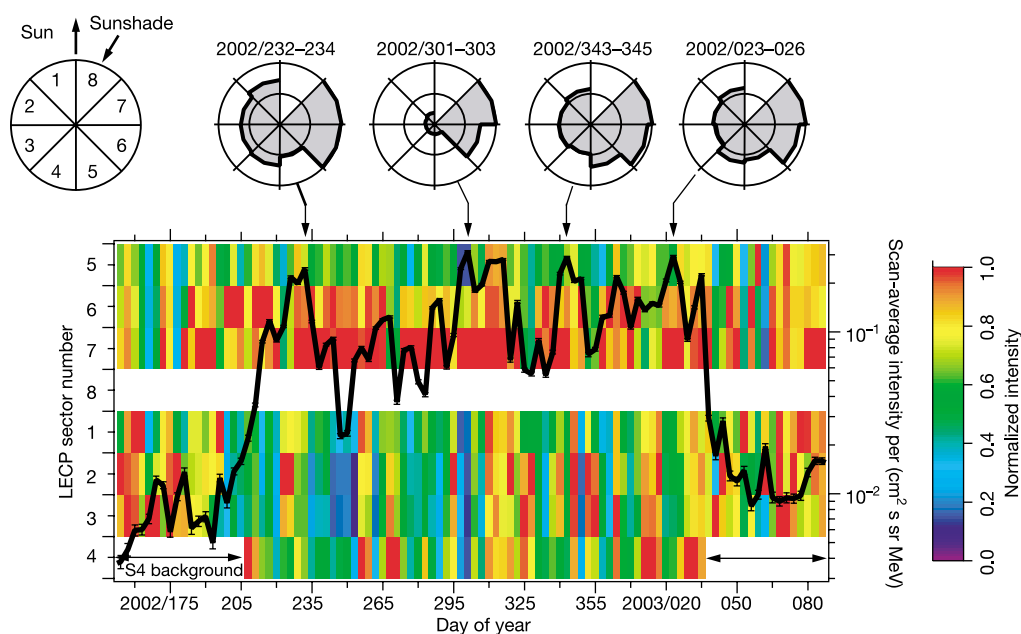


Figure 2 Data from Voyager 1 LECP for 0.57–1.78 MeV protons. Pie plots show three-day averaged normalized intensities of ~ 1 -MeV protons for four selected flux peaks; the colour scale depicts arrival directions for all the data. The black trace shows the sector-

averaged intensities. Sector 8 (see key to pie plots) contains a sunshade, so no foreground data are obtained in that direction.

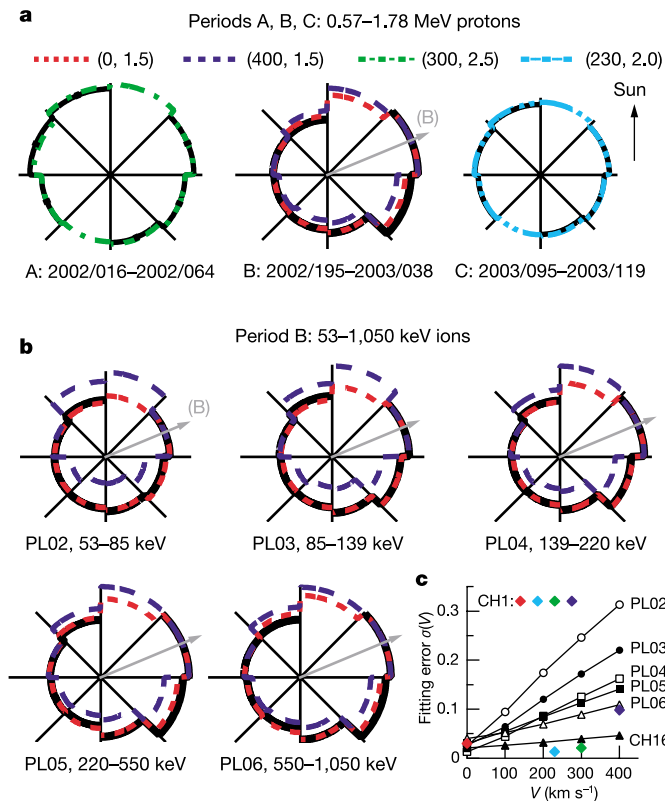


Figure 3 Anisotropy plots of time-averaged sector rates for periods A, B and C noted in Fig. 1 and solar wind velocity fits thereof. Solid black curves, data. Coloured dashed curves, models for (V, γ) cases (V in km s⁻¹). **a**, Data fits for ~1-MeV protons for periods when Voyager 1 was upstream (left), in heliosheath (middle), and upstream again (right), with best fits of ~300 km s⁻¹, 0 km s⁻¹, and ~230 km s⁻¹, respectively. **b**, At lower energies (Voyager 1 in heliosheath), model-predicted rates best-fit data (statistical error is less than the line thickness) at 0 km s⁻¹. 'B' indicates the best-fit particle streaming direction. **c**, Model-fitting errors for each channel show the degeneration of fit as velocity increases.

increases at Voyager 1 from 2002.5 to 2003.1 with no corresponding increases at Voyager 2, and the complete lack of a Forbush decrease in either ACR or GCR intensities, making an interplanetary disturbance a highly unlikely explanation. The ~20% GCR increase could be a manifestation of either additional acceleration at the termination shock at $E > 70$ MeV or a substantial gradient in GCR intensity between the heliosphere proper and the heliosheath, or both. The 'decreased modulation' scenario suggested⁹ to explain the GCR increase cannot be valid, because such a change in modulation was not seen earlier by Voyager 2 (Fig. 1c), as is generally the case for convective delays. The anisotropies are directed outward, implying that Voyager 1 is beyond the termination shock and that the composition of H, He, C and O is consistent with both ACRs and pick-up ions, as would be expected near the termination shock.

Slowing of solar wind flow must be accompanied by changes in interplanetary magnetic field as required by the 'frozen-in' magnetic-field condition¹. Only a small (~50–100%) magnetic field increase has been reported¹⁸ in mid-2002. The observation that particles are streaming ~22° off the tangential direction (Fig. 3), presumably along the interplanetary magnetic field, suggests that the field is not in the Parker spiral direction as would be expected if Voyager 1 was still in the supersonic solar wind. The interplanetary magnetic field at this latitude (34°) may be more like the Fisk field¹⁹, possibly deformed by FALTS (favoured acceleration locations at the

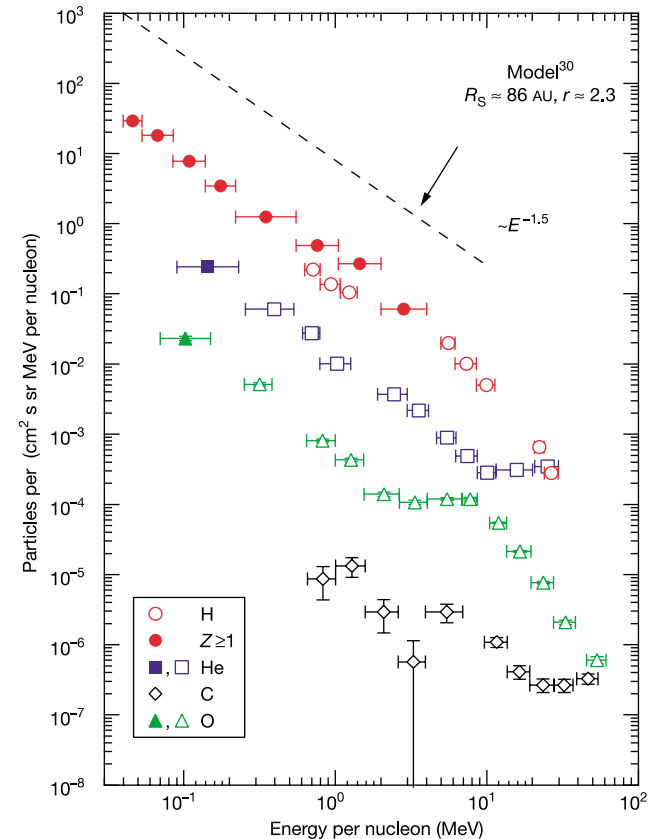


Figure 4 Composition spectra for the duration of Voyager 1's excursion into the heliosheath from LECP 2002/194 to 2003/044. At lower energies the spectrum is of the form $j = KE^{-\gamma}$, with $\gamma \approx 1.5$ for the three major species. There is no commensurate increase in the intensity of C ($C/O \approx 0.02$ at ~1 MeV). The relative abundances at ≤ 2 MeV per nucleon are consistent with the ACR H, He and O measurements made by LECP during the peak of the most recent recovery period (~1995–1999) (ref. 17 and references therein), as well as pick-up ions⁴. The higher-energy component is of modulated ACRs observed by LECP before this event. Dashed line shows model predictions³⁰.

termination shock²⁰) being more radial than the Parker field, with its strength decreasing more like R^{-2} . Hence the ~0.044-nT magnetic field²⁰ may well be consistent with the observed decrease in the solar wind velocity.

The high fluxes of pick-up ions would change the nature of the termination shock from the conventional model, reducing the upstream Mach number, hence weakening the shock, and slowing down the solar wind^{21–23} to 200–300 km s⁻¹, as suggested by our measurements. Such a strongly moderated shock has never been observed previously anywhere in the heliosphere. The closest analogues may be mass-loaded cometary shocks that produce copious quantities of accelerated ions and rarely show a definitive change in magnetic-field strength at crossing the bow shock^{24,25}. The putative termination shock may well be a new variant that will test our understanding of the overall shock formation and acceleration processes. □

Methods

We extract the solar wind velocity (V) by using the effect produced by the convection on the LECP intensity anisotropy. The Galilean transformation into the spacecraft frame is $f(\mathbf{v}) = f'(\mathbf{v} - \mathbf{V})$. The unidirectional differential intensity (which LECP measures) is $j(E, \mathbf{u}) = (v^2/m)f(\mathbf{v}) = (E/E')^2 j'(E', \mathbf{u}')$, where E is the energy and the particle velocity direction is $\mathbf{u} = \mathbf{v}/v$. It is usually assumed²⁶ that $j'(E', \mathbf{u}')$ is isotropic (independent of $\mathbf{u}'$) and that the transformation is linearized to $O(V/v)$. That approximation has been applied successfully¹⁶ to the extraction of the solar wind velocity from the LECP anisotropy

measurements on Voyager 1 and Voyager 2. However, the current Voyager 1 LECP measurements indicate significant field-aligned anisotropies strong enough to invalidate the use of the linearized transformation. Consequently, on the basis of gyrotropic weak-scattering theory²⁷, we assume an exponential distribution in pitch-cosine $j' = j_0(E) \exp[\alpha(v')\mu']$ that is convected with the solar wind²⁸. We transform it (nonlinearly) into the spacecraft frame: $j(E, \mu) = j_0(E) C^{-(k+1)} \exp[\alpha(v)(\mu - V_{\parallel}/v) C^{(n-1)/2}]$, where $C = E'/E = 1 - 2V_{\parallel}v/v^2 + V^2/v^2$, $V_{\parallel}$ is the component of the solar wind velocity along the magnetic field. We have also assumed that $V \ll v$, $j_0 \propto (E')^{-k}$ and $\alpha \propto (v')^n$. The measured spectral index is $k = 1.5$ and we find that $n \approx 0$ over the range of LECP energies. We allow for the weak coupling (backscatter) between hemispheres by setting $\alpha = \alpha_+$ for $\mu > V_{\parallel}/v$ and $\alpha = \alpha_-$ for $\mu < V_{\parallel}/v$. The LECP distributions consistently peak in sector 7, so we assign the direction of the magnetic field to its centre and normalize the intensities there (thus eliminating j_0). The normalized intensities in the remaining six sectors (sector 8 is blocked) for each LECP channel are then fitted by a least-squares minimization that varies the remaining parameters (α_+ , α_- , and V). Thus we extract the solar wind velocity from the LECP angular distributions (see Fig. 3).

We could not explain our observations using diffusion-convection (strong-scattering) theory²⁹ under the assumption that Voyager 1 did not leave the normal solar wind and magnetic field, as suggested^{29,18}. The equation describing the radial transport is $\xi_r = (3/v)(CV - \kappa_{rr}G_r)$, where ξ_r is the radial anisotropy coefficient, $C = 2(k + 1)/3$ is the (linearized) Compton-Getting factor, $G_r = \partial \ln j / \partial r$ is the radial (logarithmic) gradient, and κ_{rr} is the relevant component of the diffusion tensor. The condition for diffusion-convection equilibrium with no radial streaming in the inertial frame is $\xi_r = 0$. This implies a positive radial gradient, $\kappa_{rr}G_r = CV > 0$, with a source of particles beyond Voyager 1 in order to nullify the solar wind convection. For an approximately diagonal diffusion tensor (appropriate to the outer heliosphere), the azimuthal anisotropy is $\xi_\phi = (3/v)(-\kappa_{\phi\phi}G_\phi)$. Because LECP observes a strong azimuthal anisotropy $\xi_\phi < 0$, this implies $G_\phi > 0$. If the average magnetic field is wound in a Parker sense, $G_\phi > 0$ corresponds to gradient of increasing intensity as one moves inward along the field. This parallel streaming then implies a source of particles inside the radius of Voyager, but this is inconsistent with the positive radial gradient $G_r > 0$ demanded by the radial transport equation.

We examined the quantitative implications of the diffusion-convection equation for long-term averages of the radial anisotropy $\langle \xi_r \rangle$ and its variance $\sigma_\xi^2 = \langle \delta \xi_r^2 \rangle$. Diffusive balance implies that $\langle \xi_r \rangle = 0$, so that $\kappa_{rr}G_r = CV$, under the assumption of no significant variation in either the solar wind velocity or the diffusion coefficient. The diffusion-convection equation also implies that $\sigma_\xi = \kappa_{rr}\sigma_G$, where $\sigma_G^2 = \langle \delta G_r^2 \rangle$. Taking the ratio of the mean and root-mean-square equations, the diffusion coefficient cancels out and we find that $\sigma_G / \langle G_r \rangle = v \sigma_\xi / 3CV$. We have estimated the mean $\langle \xi_r \rangle$ and the variance σ_ξ^2 of the radial anisotropy measured directly by LECP Channel 1 from the ratio of Sector 1 to Sector 5 (see Fig. 2). From $\xi_r = (S1 - S5)/(S1 + S5) \approx (S1/S5 - 1)/2$, we find during period B that $\langle \xi_r \rangle \approx -0.015$ and $\sigma_\xi = 0.15$. Using the observed $k = 1.5$ we have $C = 5/3$. Using 350 km s^{-1} for the normal solar wind velocity and $14,000 \text{ km s}^{-1}$ for the 1-MeV protons in channel 1, we conclude that $v \sigma_\xi / 3CV = 1.2$. This value implies an abnormally disturbed region with $\sigma_G \approx \langle G_r \rangle$. We can estimate σ_G itself in the diffusion-convection context by assuming that the quasi-sinusoidal non-dispersive LECP intensity variations δj with period T were caused by spatial structures with instantaneous gradients (δG_r) being convected with velocity V over the spacecraft so that $V \delta G_r = \partial \ln j / \partial t$. There are about a dozen large $(\Delta \ln j > 1)$ sinusoidal variations in LECP channel 1 over the 208 days during period B ($T \approx 16$ days). Therefore $V \sigma_G > 0.29 \text{ day}^{-1}$, so for $V = 350 \text{ km s}^{-1}$, we have the estimate $\sigma_G = 1.44 \text{ AU}^{-1}$. Since we previously found that $\langle G_r \rangle \approx \sigma_G$, we obtain a time-averaged radial gradient $\langle G_r \rangle \approx 1 \text{ AU}^{-1}$ that (along with its variation σ_G) is orders of magnitude bigger than gradients usually deduced for the outer heliosphere. We consider it unreasonable that such a configuration could endure there for half a year.

Received 27 May; accepted 18 September 2003; doi:10.1038/nature02068.

- Parker, E. N. The stellar-wind regions. *Astrophys. J.* **134**, 20–27 (1961).
- Stone, E. C. News from the edge of interstellar space. *Science* **293**, 55–56 (2001).
- Fisk, L. A., Kozlovsky, B. & Ramaty, R. An interpretation of the observed oxygen and nitrogen enhancements in low-energy cosmic rays. *Astrophys. J.* **190**, L35–L37 (1974).
- Gloeckler, G. & Geiss, J. in *AIP Conf. Proc. 598 Joint SOHO-ACE Workshop 2001* (ed. Wimmer-Schweinbruber, R.) 281–289 (AIP, Melville, NY, 2001).
- Pesses, M. E., Jokipii, J. R. & Eichler, D. Cosmic ray drift, shock acceleration, and the anomalous component of cosmic rays. *Astrophys. J.* **246**, L85–L88 (1981).
- Steenberg, C. D. & Moraal, H. Form of the anomalous cosmic ray spectrum at the solar wind termination shock. *J. Geophys. Res.* **104**, 24879–24884 (1999).
- Fichtner, H. Anomalous cosmic rays: Messengers from the outer heliosphere. *Space Sci. Rev.* **95**, 639–754 (2001).
- Krimigis, S. M. *et al.* The Low Energy Charged Particle (LECP) experiment on the Voyager spacecraft. *Space Sci. Rev.* **21**, 329–354 (1977).
- McDonald, F. B. *et al.* Enhancements of energetic particles near the heliospheric termination shock. *Nature* **426**, 48–51 (2003).
- Decker, R. B., Krimigis, S. M., McNutt, R. L., Hamilton, D. C. & Collier, R. M. Latitude associated differences in the low energy charged particle activity at Voyagers 1 and 2 during 1991 to early 1994. *Space Sci. Rev.* **72**, 347–352 (1995).
- Decker, R. B., Krimigis, S. M., Roelof, E. C. & Hill, M. E. Angular distributions and energy spectra of low-energy ions observed by Voyager 1 at 85–88 AU. *Geophys. Res. Abstr.* **5**, 03301 (2003).
- Roelof, E. C., Simnett, G. M., Sanderson, T. R. & Kunow, H. Corotating interaction regions at high latitudes. *Space Sci. Rev.* **89**, 225–233 (1999).
- Decker, R. B., Roelof, E. C. & Krimigis, S. M. in *Acceleration and Transport of Energetic Particles in the Heliosphere* (eds Mewaldt, R. A., Zurbuchen, T. H. & Cummings, A. C.) *AIP Conf. Proc.* **528**, 161–165 (AIP, Melville, NY, 2000).

- Decker, R. B. The role of magnetic loops in particle acceleration at nearly perpendicular shocks. *J. Geophys. Res.* **98**, 33–46 (1993).
- Sarris, E. T. & Krimigis, S. M. Quasi-perpendicular shock acceleration of ions to ~ 200 MeV and electrons to ~ 2 MeV observed by Voyager-2. *Astrophys. J.* **298**, 676–683 (1985).
- Kane, M., Decker, R. B., Mauk, B. H. & Krimigis, S. M. The solar wind velocity determined from Voyager 1 and 2: Low Energy Charged Particle measurements in the outer heliosphere. *J. Geophys. Res.* **103**, 267–276 (1998).
- Hill, M. E., Hamilton, D. C. & Krimigis, S. M. Evolution of anomalous cosmic-ray oxygen and helium energy spectra during the solar cycle 22 recovery phase in the outer heliosphere. *Astrophys. J.* **572**, L169–L172 (2002).
- Burlaga, L. F. *et al.* Search for the heliosheath with Voyager 1 magnetic field measurements. *Geophys. Res. Lett.* (in the press).
- Fisk, L. A. Motion of the footpoints of heliospheric magnetic field lines at the sun: Implications for recurrent energetic particle events at high heliographic latitudes. *J. Geophys. Res.* **101**, 15547–15553 (1996).
- Schwadron, N. A. & McComas, D. J. Heliospheric “FALTS”: Favored acceleration locations at the termination shock. *Geophys. Res. Lett.* **30**, doi: 10.1029/2002GL016499 (2003).
- Izmodenov, V. V., Gloeckler, G. & Malama, V. When will Voyager 1 and 2 cross the termination shock? *Geophys. Res. Lett.* **30**, 3–14 (2003).
- Fahr, H. J. & Rucinski, D. Neutral interstellar gas atoms reducing the solar wind Mach number and velocity. *Astron. Astrophys.* **350**, 1071–1078 (1999).
- Fahr, H. J., Kausch, T. & Scherer, H. A five fluid hydrodynamic approach to model the solar system-interstellar medium interaction. *Astron. Astrophys.* **357**, 268–282 (2000).
- Neugebauer, M. Spacecraft observations of the interaction of active comets with the solar wind. *Rev. Geophys.* **28**, 231–252 (1990).
- Cargill, P. J., Hizanidis, K. & Papadopoulos, K. in *Cometary and Solar Plasma Physics* (ed. Buti, B.) (World Scientific, New York, 1988).
- Gleeson, L. J. & Axford, W. I. The Compton-Getting effect. *Astrophys. Space Sci.* **2**, 431–437 (1968).
- Roelof, E. C. in *Lectures in High Energy Astrophysics* (eds Ogelman, H. & Wayland, J. R.) Ch. VII (NASA SP-199, 1969).
- Zwicky, R. D. & Roelof, E. C. Interplanetary propagation of < 1 MeV protons in non-impulsive energetic particle events. *J. Geophys. Res.* **86**, 5449 (1981).
- Parker, E. N. The passage of energetic charged particles through interplanetary space. *Planet. Space Sci.* **13**, 9–49 (1965).
- Le Roux, J. A., Fichtner, H., Zank, G. P. & Ptuskin, V. S. Self-consistent injection and acceleration of pickup ions at the solar wind termination shock. *Geophys. Res. Lett.* **27**, 2873–2876 (2000).

Acknowledgements This work was supported by a NASA grant to the Johns Hopkins University and by subcontracts at the University of Maryland and Fundamental Technologies.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.M.K. (tom.krimigis@jhuapl.edu).

Enhancements of energetic particles near the heliospheric termination shock

Frank B. McDonald¹, Edward C. Stone³, Alan C. Cummings³, Bryant Heikkilä², Nand Lal² & William R. Webber⁴

¹Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, and ²NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

³California Institute of Technology, Pasadena, California 91125, USA

⁴Department of Physics and Astronomy, New Mexico State University, Las Cruces, New Mexico 88003, USA

The spacecraft Voyager 1 is at a distance greater than 85 AU from the Sun, in the vicinity of the termination shock that marks the abrupt slowing of the supersonic solar wind and the beginning of the extended and unexplored distant heliosphere^{1,2}. This shock is expected to accelerate ‘anomalous cosmic rays’³, as well as to re-accelerate Galactic cosmic rays⁵ and low-energy particles from the inner Solar System⁴. Here we report a significant increase in the numbers of energetic ions and electrons that persisted for seven months beginning in mid-2002. This increase differs from any previously observed in that there was a simultaneous increase in Galactic cosmic ray ions and electrons, anomalous

cosmic rays and low-energy ions. The low-intensity level and spectral energy distribution of the anomalous cosmic rays, however, indicates that Voyager 1 still has not reached the termination shock. Rather, the observed increase is an expected precursor event. We argue that the radial anisotropy of the cosmic rays is expected to be small in the foreshock region, as is observed.

The energetic particle data reported here are from the Voyager 1 and 2 cosmic ray subsystems (CRS)⁶. Starting in mid-2002, increases of low-energy electrons and ions were observed that persisted for some six months (Fig. 1). The time histories of these 2.5-MeV H and 10-MeV electrons are similar both at the beginning of the event and for the shorter-term variations. Over this period the 20–69 MeV per nucleon anomalous cosmic ray (ACR) He intensity increased by a factor of 1.5 while the Pen L rate (70- to ~300-MeV H, He > 70 MeV per nucleon) increased by 15%. The short-term variations of these two components are similar and generally resemble those of the MeV ions and electrons. In the past, increases in low-energy ions were associated with the passage of transients in the interplanetary medium that results in a decrease in galactic cosmic rays (GCRs) and ACRs. There is no evidence for the passage of even moderate interplanetary disturbances until decimal year 2003.09, that is, 2 February 2003.

Over this period from 2000.8 to 2003.2 a series of six well-defined but modest enhancements of MeV protons at Voyager 2 occurred quasi-periodically every ~140 days, persisted on average for 0.3 yr (~4 solar rotations) (Fig. 2), and were accompanied by transient decreases in GCRs and ACRs. For the first three events the Voyager 1 peak intensity was on average 40% of that at Voyager 2, indicating a

mean radial intensity gradient of -5.5% per AU. In contrast, the Voyager 1 event beginning in mid-2002 persisted for more than six months with intensities 20 to 500 times greater than at Voyager 2. Such a large disparity has not previously been observed in the outer heliosphere.

At Voyager 1 the cosmic ray decreases associated with events I and III (Fig. 2) are comparable to those at Voyager 2. However, there is no evidence of any significant interplanetary disturbance in the Voyager 1 cosmic ray data in the 9-month period from 2002.3 until 2003.09. At that time, the 2.5-MeV H intensity decreased by a factor of ~50 in two steps in coincidence with similar but smaller decreases in the Pen L rate (Fig. 1, 2), indicating the passage of an interplanetary disturbance which was probably produced by the increased solar activity in July 2002. For 2.5-MeV H this reduced level is still higher than observed by Voyager 1 during events I–III.

The large enhancement in 2002 is apparent at low energies in the H and He spectra in Fig. 3. Although the Voyager 1 intensities at low energies are similar to the predicted source spectra⁷ of ACRs at the shock where they are accelerated, the observed ACR He intensity at intermediate energies (10–20 MeV per nucleon) is <0.1 of the predicted source spectrum and less than was observed⁷ at solar minimum (year 1998/day 1 to 1999/182) when Voyager 1 was at 69–74 AU. Given current models of ACR shock acceleration, this residual modulation would indicate that Voyager 1 did not encounter the shock.

The four CRS Low-Energy Telescopes provide measurements of the directional flow of ions^{6,8}. The period from 2002/195 to 2003/38 was dominated by episodes of unusually large azimuthal streaming, consistent with strong flows away from the sunward direction along the magnetic field. As the field direction varied inward and outward from its nominal azimuthal direction, there were significant inward and outward radial components of the field-aligned flows. Outward and inward radial flows differed from zero by $>1\sigma$ on 55 and 56 days, respectively, and by $<1\sigma$ on 56 and 42 days, respectively, out of a total of 209 days. Although there were large flows on individual days, the average radial anisotropy over the entire 209-day period was only 0.006 ± 0.011 , much smaller than the expected convective anisotropy of 0.09 for a 400 km s^{-1} solar wind. If the field-aligned flows were purely azimuthal throughout the entire period, this would indicate a small convective streaming as would result from a heliosheath wind speed of $<100 \text{ km s}^{-1}$, as suggested in ref. 9. However, the flow directions were highly variable and the radial component was usually much larger than this average, indicating that the average was dominated by the difference of large positive and negative radial components of the field-aligned flows rather than by the much smaller convective flow.

The anisotropy just upstream of the shock is determined by the requirements that the anisotropy and intensity be continuous across the shock^{5,10,11}. This requires that the average radial upstream anisotropy (within ~10 AU of the termination shock) equal that downstream, which is determined by the subsonic flow in the heliosheath¹² and should be ≤ 0.02 . Thus, the average upstream radial anisotropy should be the same as in the subsonic heliosheath even though the upstream wind is faster, meaning that the larger convective anisotropy of 0.09 must be substantially balanced by inward diffusion¹⁰. This is consistent with magnetic-field observations that did not show the increased magnetic-field strength expected if Voyager 1 was in the heliosheath during this period¹³.

Over the past 17 years the largest increase in 2.5-MeV ions (September 1991, 46 AU) was about three times larger than observed in 2002. If the recent increase was treated as being of solar origin, then applying the radial gradient correction of -5.5% per AU between 86 and 46 AU would mean that the 2002 Voyager 1 increase was three times larger than that of 1991. However, the 1991 event produced the largest neutron monitor Forbush decrease ever observed at 1 AU, a large transient decrease at Voyager 1, and long-lasting radio emission from the region beyond the termination

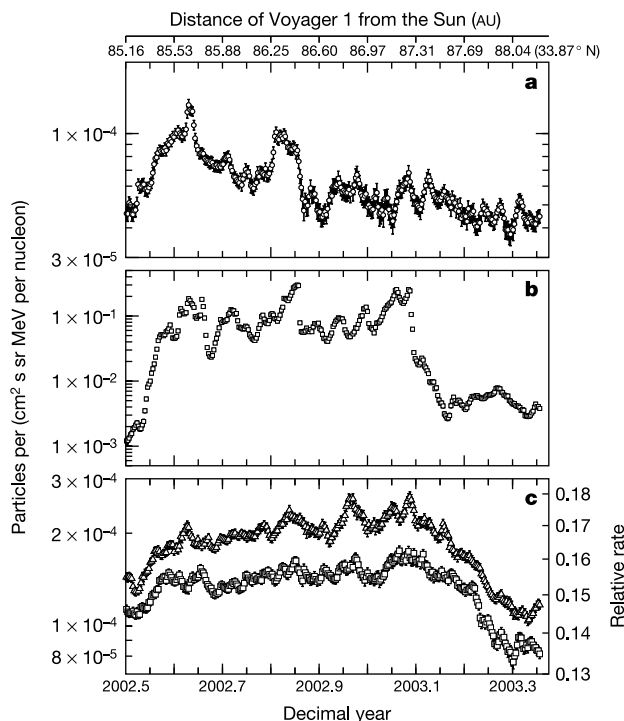


Figure 1 Time histories (5-day moving averages) of Voyager 1 cosmic ray intensities. **a**, 6–14-MeV electrons. **b**, 2–3 MeV H. **c**, 27–69 MeV per nucleon ACR He⁺ (open squares) and the Pen L relative rate (H ions 70 < E < 300 MeV, and $Z > 1$ ions $E > 70$ MeV per nucleon) (open triangles). The data are from the Voyager CRS instruments, which consist of seven multi-element solid-state energetic particle telescopes covering electrons (2–160 MeV), H and He (1.8–500 MeV per nucleon) and heavier nuclei through to Fe. Over this period, Voyager 1 is at a mean heliolatitude of 33.87° N.

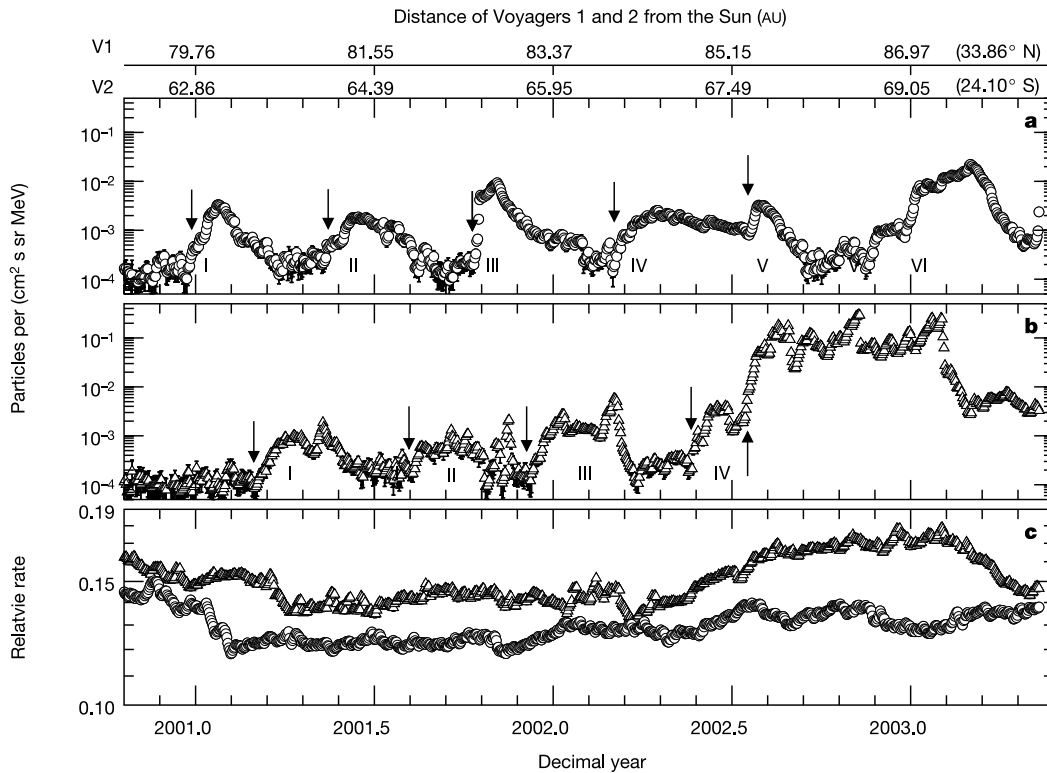


Figure 2 Time histories (5-day moving averages) of Voyager 1 (V1) and Voyager 2 (V2) cosmic-ray intensities. **a**, Voyager 2, 2–3 MeV H. **b**, Voyager 1, 2–3 MeV H. **c**, The Voyager 1 Pen L rate (open triangles) and Voyager 2 Pen L rate (open circles). Five of the Voyager 2 events were closely associated with increases of 100 km s^{-1} in the solar wind

speed at Voyager 2, which can be identified with specific periods of activity occurring some 5–6 months earlier at the Sun. Events I–IV occurred at Voyager 1 after a delay of ~ 0.2 years corresponding to a solar wind speed of 420 km s^{-1} . Over this period Voyager 1 was at a mean heliographic latitude of 33.86° N and Voyager 2 was at 24.1° S .

shock^{14–16}. There have been similar MeV electron events of solar/interplanetary origin observed at Voyager 2 (30 AU) over the 1989–1991 period but they had large negative radial gradients and were barely detectable at Voyager 1, then only some 10 AU further out¹⁷.

The reduction in the modulation level in the distant heliosphere

in mid-2002 is evident from the simultaneous increase of GCRs, ACRs, MeV ions and electrons. We suggest that the 2.5–14-MeV electrons are of galactic origin that have been reaccelerated at the termination shock. This would indicate that particles of very low rigidity travelled from the termination shock to Voyager 1 at this time and the 2.5-MeV ions observed after 2002.54 are the solar/interplanetary ions with also a possible significant contribution from interstellar pick-up ions that have been preferentially accelerated¹⁸ in interplanetary space and then reaccelerated at the termination shock. The intensity decrease in GCRs, ACRs and MeV ions and electrons in early February 2003 is exactly what would be produced by the passage of a small to moderate interplanetary disturbance through a pre-existing population of these components. The earlier commonality of short-term changes would be their response to much smaller transients.

There are two major difficulties with this interpretation: (1) the peak of the ACR He intensity at this time is $\sim 25 \text{ MeV per nucleon}$ (see Fig. 3) compared to a peak energy of 6.5 MeV per nucleon during the 1998 solar minimum, suggesting there is still significant modulation between Voyager 1 and the termination shock. It may be that the low-energy component is accelerated in a localized region of the termination shock that is nearby, while the higher-energy ACRs are accelerated in a more distant larger-scale region^{12,19}. (2) The periods of very large anisotropies in MeV H ion data are predominantly field-aligned with flow in the direction away from the Sun⁸. If the shock is not the source of the energetic ions, significant interplanetary acceleration must have occurred²⁰.

Our understanding of these unusual enhancements of ions and electrons beyond 85 AU should increase as Voyager 1 continues to approach the termination shock and as the current-sheet tilt continues to evolve with declining solar activity over the next several years. □

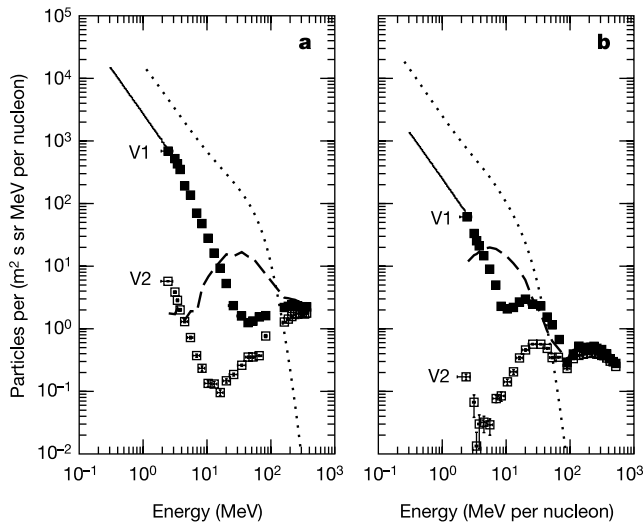


Figure 3 Voyager 1 and Voyager 2 energy spectra (2002/209–2002/364). **a**, Hydrogen and **b**, helium spectra at Voyager 1 and Voyager 2 during the intensity enhancement, along with the solar minimum spectra⁷ observed by Voyager 1 in 1998/1–1999/182 (dashed lines) and the predicted shock source spectra⁷ for a weak shock (dotted lines). Anomalous cosmic rays are predominantly singly charged, moderate-velocity, high-rigidity ions that have their origin as interstellar neutrals which have been ionized in interplanetary space, convected outward by the solar wind and accelerated by the termination shock.

Received 28 May; accepted 14 September 2003; doi:10.1038/nature02066.

1. Parker, E. N. *Interplanetary Dynamical Processes* 115–128 (Interscience Division, John Wiley and Sons, New York, 1963).
2. Holzer, T. E. Interaction between the solar wind and the interstellar medium. *Annu. Rev. Astron. Astrophys.* **270**, 199–234 (1989).
3. Pesses, M. E., Jokipii, J. R. & Eichler, D. Cosmic ray drift, shock wave acceleration and the anomalous component of cosmic rays. *Astrophys. J.* **246**, L85–L88 (1981).
4. Mewaldt, R. Solar and interplanetary particles reaccelerated at the solar wind termination shock. *Proc. 24th ICRC (Rome)* **3**, 804–807 (1995).
5. Jokipii, J. R., Kota, J. & Merenyi, E. The gradient of galactic cosmic rays at the solar wind termination shock. *Astrophys. J.* **405**, 782–786 (1993).
6. Stone, E. C. *et al.* Cosmic ray investigations for the Voyager missions: Energetic particle studies in the outer heliosphere—and beyond. *Space Sci. Rev.* **21**, 355–376 (1977).
7. Cummings, A. C., Stone, E. C. & Steenberg, C. D. Composition of anomalous cosmic rays and other interplanetary ions. *Astrophys. J.* **578**, 194–210 (2002).
8. Cummings, A. C. *et al.* Voyager 1 observations of the anisotropies of enhanced MeV ion fluxes at 85AU. *Proc. 28th ICRC (Tsukuba)* **7**, 3777–3780 (2003).
9. Krimigis, S. M. *et al.* Voyager 1 left the solar wind at a distance of ~85AU from the Sun. *Nature* **426**, 45–48 (2003).
10. Jokipii, J. R., Giacalone, J. & Kota, J. Anomalous cosmic ray and plasma signatures of a termination shock crossing. *Geophys. Res. Lett.* (submitted).
11. Chalov, S. V. & Fahr, H. J. Pick-up ion acceleration at the termination shock and the post-shock pick-up ion energy distribution. *Astron. Astrophys.* **360**, 381–390 (2000).
12. Jokipii, J. R. & Giacalone, J. Anomalous cosmic rays at a termination shock crossing. *Proc. 28th ICRC (Tsukuba)* **7**, 3753–3756 (2003).
13. Burlaga, L. F. *et al.* Search for the heliosheath with Voyager 1 magnetic field measurements. *Geophys. Res. Lett.* (in the press).
14. Van Allen, J. A. & Fillius, R. W. Propagation of large Forbush decreases in cosmic ray intensity past the Earth, Pioneer 11 at 34AU and Pioneer 10 at 53AU. *Geophys. Res. Lett.* **19**, 1423–1426 (1992).
15. Webber, W. R. & Lockwood, J. A. Giant transient decreases of cosmic rays in the outer heliosphere in September 1991. *J. Geophys. Res.* **98**, 7821–7826 (1991).
16. Gurnett, D. A., Kurth, W. S., Allendorf, S. C. & Poynter, R. L. Radio emission from the heliopause triggered by an interplanetary shock. *Science* **262**, 199–203 (1993).
17. Lukasiak, A. *et al.* The Voyager electron telescope—A status report. *Proc. 26th ICRC (Salt Lake City)* **3**, 65–68 (1999).
18. Gloeckler, G. *et al.* Acceleration of interstellar pickup ions in the disturbed solar wind observed on Ulysses. *J. Geophys. Res.* **99**, 17637–17643 (1994).
19. Schwadron, N. A. & McComas, D. J. Heliospheric “Falts”: Favored Acceleration Locations at the Termination Shock. *Geophys. Res. Lett.* **30**, doi:10.1029/2002GL016499 (2003).
20. Christon, S. P. & Stone, E. C. Latitude variation of recurrent MeV energy proton flux enhancements in the radial range 11–20AU and possible correlation with solar coronal hole dynamics. *Geophys. Res. Lett.* **12**, 109–112 (1985).

Acknowledgements We thank J. D. Richardson and colleagues for making the Voyager 2 solar wind data available at (<http://web.mit.edu/afs/athena/org/s/space/www/voyager.html>). We also thank J. R. Jokipii, F. C. Jones, H. Moraal, V. Ptuskin and R. Caballero-Lopez for discussions on shock acceleration processes and particle flow in the vicinity of the acceleration region.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.B.M. (fm27@umail.umd.edu).

The birth of a quasiparticle in silicon observed in time–frequency space

Muneaki Hase¹, Masahiro Kitajima¹, Anca Monia Constantinescu² & Hrvoje Petek²

¹Materials Engineering Laboratory, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan

²Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

The concept of quasiparticles in solid-state physics is an extremely powerful tool for describing complex many-body phenomena in terms of single-particle excitations¹. Introducing a simple particle, such as an electron, hole or phonon, deforms a many-body system through its interactions with other particles. In this way, the added particle is ‘dressed’ or ‘renormalized’ by a self-energy cloud that describes the response of the many-body system, so forming a new entity—the quasiparticle. Using ultra-

fast laser techniques, it is possible to impulsively generate bare particles and observe their subsequent dressing by the many-body interactions (that is, quasiparticle formation) on the time and energy scales governed by the Heisenberg uncertainty principle². Here we describe the coherent response of silicon to excitation with a 10-femtosecond (10^{-14} s) laser pulse. The optical pulse interacts with the sample by way of the complex second-order nonlinear susceptibility to generate a force on the lattice driving coherent phonon excitation. Transforming the transient reflectivity signal into frequency–time space reveals interference effects leading to the coherent phonon generation and subsequent dressing of the phonon by electron–hole pair excitations.

Silicon is a model for the fundamental electronic and mechanical properties of Group IV and III–V semiconductors and the basic material for electronic device technology. The interaction between electrons and phonons via the Coulomb force defines many of the physical properties of Si, such as the electrical conductivity: real and virtual scattering processes transfer carriers between different momentum states and renormalize their mass¹. Much of our knowledge of the electron–phonon interaction in Si is limited to transport measurements at near-equilibrium conditions; but as the electronic device dimensions become comparable to carrier scattering lengths, the fundamental quantum mechanical carrier–carrier and carrier–lattice interactions increasingly determine their properties³. Measurements of the carrier relaxation in silicon by optical techniques have been restricted mainly to transient changes of index of refraction following excitation by intense laser pulses^{4–7}, because the indirect bandgap of Si obviates more specific probes. It has been established that the quantum mechanical phase of the primary photo-excited electron–hole pairs (momentum scattering) decays in $\ll 100$ fs, whereas charge carriers equilibrate with the lattice on a timescale of 200–300 fs (refs 5–9).

Here we explore the coherent response of the coupled Si carrier–lattice system to excitation by a 10-fs laser pulse. The transient anisotropic reflectivity signal is measured and transformed into time–frequency space, revealing the dynamics of formation of the zone-centre (zero-momentum) coherent longitudinal optical (LO) phonon at 15.3 THz (ref. 6), and its subsequent dressing by interaction with the photogenerated carrier plasma. The birth of a quasiparticle, the coherent phonon, is observed in the quantum kinetic regime, where energy–time uncertainty and phase memory stored by the coherent polarization make the carrier–light, carrier–carrier and carrier–phonon interactions non-local in time².

The anisotropic transient reflectivity of n-doped Si(001) (doping concentration $\sim 1 \times 10^{15} \text{ cm}^{-3}$) under ambient conditions is measured using the electro-optic sampling method¹⁰. Nearly collinear, pump and probe pulses (10-fs pulse duration; 406-nm (3.05-eV) wavelength; 90-MHz repetition rate) are focused to a 10- μm spot on the sample with an average angle of 5° from the surface normal. The pump pulse with 40 mW average power generates a carrier density of $4 \times 10^{19} \text{ cm}^{-3}$; the probe power is ≤ 2 mW. Polarization of the two beams is linear with a mutual angle of 45° , as shown in the insets of Fig. 1. After reflecting from the sample, the probe beam is analysed into polarization components parallel and perpendicular to that of the pump, $R_{\parallel}$ and $R_{\perp}$, and detected with photodiodes. The resulting photocurrents are subtracted, and their difference $\Delta R_{\text{eo}}(\tau) = R_{\perp}(\tau) - R_{\parallel}(\tau)$ recorded as a function of the pump–probe delay τ .

The excitation light interacting with the sample generates non-linear polarization $P_j^{(2)}(\tau)$ normal to the surface by the process of optical rectification:

$$P_j^{(2)}(\tau) = \chi_{jkl}^{(2)}(\tau - t) E_k(t) E_l(t) + \int_{-\infty}^{\tau} J_j(t') dt' \quad (1)$$

where $\chi_{jkl}^{(2)}(\tau - t)$ is the complex, second-order nonlinear susceptibility, $E_{k,l}(t)$ are vectorial components of the incident electric fields,

and $J_j(t')$ is the macroscopic photo-Dember current arising from differential electron/hole mobilities^{11–13}. As the bulk dipole contribution to $\chi_{jkl}^{(2)}(\tau - t)$ is zero by symmetry, the bulk quadrupole and surface dipole terms probably dominate the electro-optic response¹⁴. The optical rectification or inverse electro-optic response described by $\chi_{jkl}^{(2)}\{0; -\omega, \omega'\}$ generates coherent polarization at the beat frequency ($\omega - \omega' \approx 0$) by the difference frequency mixing of components of the carrier wave (ω, ω') (refs 11–13). Because under resonant conditions the optical rectification involves

electron–hole pair excitation with finite phase and energy relaxation times, the $\chi_{jkl}^{(2)}(\tau - t)$ response depends on the delay τ . The surface-normal [001] component of the induced nonlinear polarization modulates the sample reflectivity via the electro-optic effect¹⁰:

$$\frac{R_{\perp}(\tau) - R_{\parallel}(\tau)}{R_0} = \frac{\Delta R_{eo}(\tau)}{R_0} \approx r P_{001}^{(2)}(\tau) \quad (2)$$

where the coefficient r represents the bulk quadrupole and surface dipole electro-optic response. Because in the electro-optic sampling

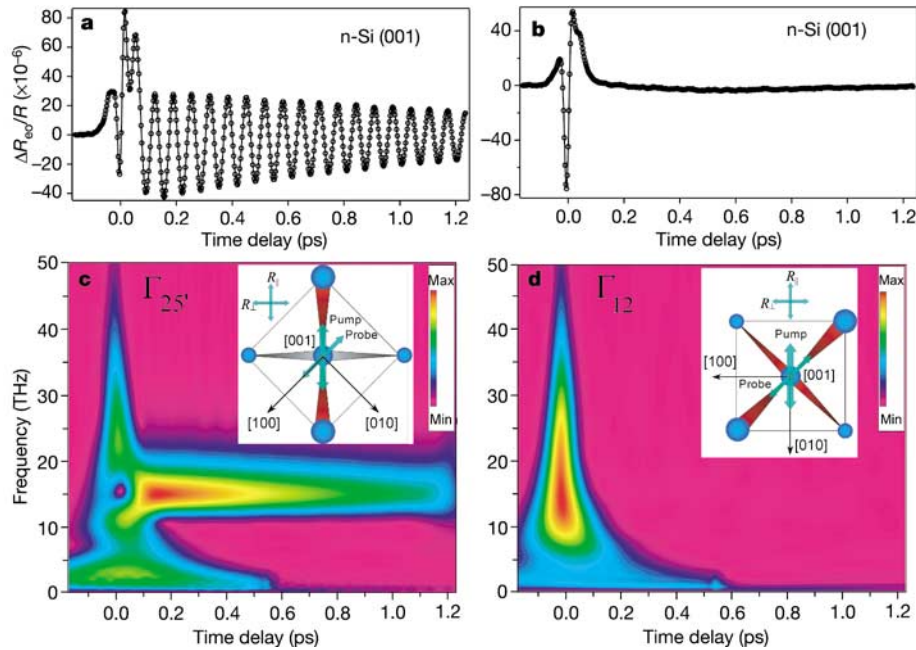


Figure 1 Transient electro-optic reflectivity signals for Si(001), and their continuous wavelet transforms. **a**, **b**, Transient electro-optic reflectivity signal for Si(001) in $\Gamma_{25'}$ (**a**) and Γ_{12} (**b**) geometry. **c**, **d**, Continuous wavelet transforms (CWT chronograms) of the data in **a** and **b**, respectively. Left insets in **c** and **d** define the polarization of the laser beams relative to the crystalline axes of the sample in the $\Gamma_{25'}$ and Γ_{12} geometries. Blue spheres indicate Si atoms in the diamond lattice with size denoting upward or downward

disposition with respect to the central atom. Excited or quiescent bonds are represented by red and grey. The signal in **a** and **b** is derived from the difference in $R_{\parallel}$ and $R_{\perp}$, the orthogonal polarization components of the probe beam. The colour scale indicates the signal amplitude, red being the maximum. Three-dimensional CWT animations can be viewed at http://www.phyast.pitt.edu/People/Faculty/H_Petek/3Dwavelet.

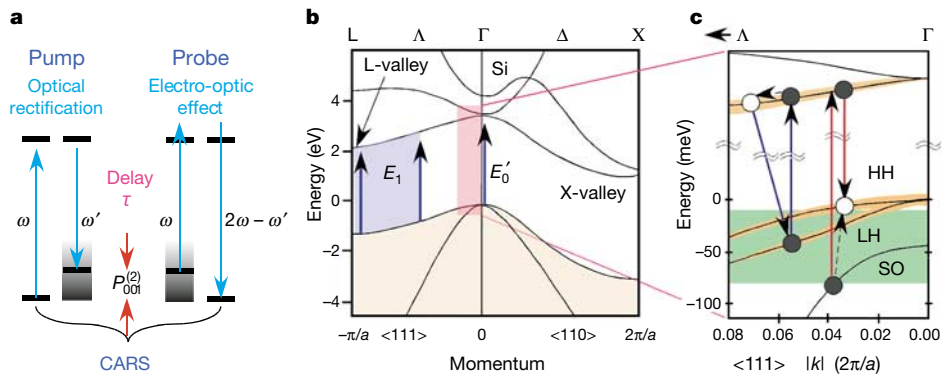


Figure 2 The excitation processes that give rise to the electro-optic reflectivity signal. **a**, The excitation scheme for electro-optic sampling. The pump pulse generates coherent polarization $P_{001}^{(2)}$ at a frequency $\omega - \omega'$. The delayed probe beam scatters from the polarization, generating light at frequency $2\omega - \omega'$. The cascaded interaction corresponds to coherent anti-Stokes Raman scattering (CARS). **b**, The band structure of Si for momenta in the Λ and Δ directions. Photoabsorption at 3.05 eV occurs near the E'_0 and E_1 critical points, corresponding to transitions near the Brillouin zone centre (Γ point) and a range of momenta along the Δ line, indicated by blue arrows and shading.

c, Electronic Raman processes that can contribute to the nonlinear polarization. Intraband scattering occurs mainly within the heavy-hole (HH) valence band, and the X- and L-valleys (blue arrows) in the conduction band. Interband scattering involves transitions among the heavy-hole (HH), light-hole (LH) and split-off (SO) bands (SO $\rightarrow$ HH excitation is indicated by red arrows). Interaction between the SO $\rightarrow$ HH continuum and the LO phonon (excitation energy indicated by green shading) contributes to the renormalization of the phonon frequency. The yellow shading conveys the relaxation of electrons towards conduction band minima and holes towards the valence band maximum.

mode the longitudinal polarization $P_{001}^{(2)}(\tau)$ modulates $R_{\perp}$ and $R_{\parallel}$ with the opposite phase, the difference $\Delta R_{\text{eo}}(\tau)$ records only the dynamics of $P_{001}^{(2)}(\tau)$, while the much larger signal from the isotropic change in reflectivity due to the incoherent carrier and lattice dynamics is eliminated¹⁰. The cascaded sequence of excitation and probing can be described by a third-order susceptibility tensor $\chi^{(3)}(\omega''; \omega', -\omega, \omega') \approx \chi^{(2)}(0; -\omega, \omega') \cdot \chi^{(2)}(\omega''; \omega', 0)$ corresponding to the coherent Stokes and anti-Stokes Raman scattering (CSRS and CARS; Fig. 2a), as long as the coherent polarization exists in the interval between the pulses¹⁵. Other fully resonant intrinsic $\chi^{(3)}$ processes, such as induced birefringence, which involve similar excitations to the cascaded processes, can also contribute to the signal.

Transient reflectivity is measured with the [110] crystalline axis of Si oriented both parallel with, and at 45° to, the pump polarization to explore the $\Gamma_{25'}$ and Γ_{12} symmetry response, respectively (see Fig. 1). The $\Delta R_{\text{eo}}(\tau)/R_0$ measurement for the $\Gamma_{25'}$ geometry in Fig. 1a is composed of an aperiodic component near zero delay followed by a coherent LO phonon oscillation. The latter component can be fitted by $A_0 \exp(-\tau/1.30) \cos[(15.24\tau + 0.016\tau^2 + 0.064)2\pi]$, where the parameters in order of appearance are the dephasing time in picoseconds, the oscillation frequency in THz, chirp, and phase shift. The departure from the LO phonon dephasing time of 3.5 ps and frequency of 15.60 THz of intrinsic Si reflects the coherent phonon self-energy, which depends on the excitation density¹⁶. By contrast, the Γ_{12} response in Fig. 1b consists of a half-cycle transient reminiscent of the aperiodic component in the $\Gamma_{25'}$ response.

To glean more insight into the early time dynamics, Fig. 1c and d shows the $\Delta R_{\text{eo}}(\tau)/R_0$ signals converted into time-dependent spectral amplitudes (chronograms) by continuous wavelet transform

(CWT; <http://www.vallen.de/wavelet>). The CWT decomposes the signal into an ultra-broadband (d.c. to >70 THz) response straddling zero delay in both geometries, and the LO phonon response at ~15.3 THz present only in the $\Gamma_{25'}$ geometry. The fast response is the coherent electronic coupling of the pump and probe fields via the nonlinear susceptibility of the sample commonly referred to as the 'coherent artefact'. This is a misnomer, because the coherent electronic coupling in absorbing matter, in addition to instantaneous response, also includes specific microscopic electronic excitations and scattering processes with finite dephasing times¹⁷. The real and virtual processes associated with near-resonant photo-excitation across the direct bandgap of Si, which contribute to $\chi^{(2)}$ and $\chi^{(3)}$, both modulate the anisotropic reflectivity and generate the driving force for the coherent phonon excitation¹⁸. The data in Fig. 1a and c resolve the build-up of the electronic force and the ensuing response of the lattice that constitute the electron–phonon interaction in solid-state materials.

Processes contributing to $\chi^{(2)}$ and $\chi^{(3)}$ are photon absorption, and resonant electronic and vibrational Raman scattering. According to Fig. 2b, 3.05-eV light is near-resonant with the E'_0 and the more intense E_1 critical points at 3.320 eV and 3.396 eV, which correspond to the excitation electron–hole pairs with near-zero momentum (Γ point), and for a range of momenta between $(2\pi/a)\frac{1}{4}, \frac{1}{4}, \frac{1}{4}$ and $(2\pi/a)\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$ along the Λ line, respectively¹⁹. Moreover, for the alignment of the pump polarization in the $\Gamma_{25'}$ geometry (Fig. 1c inset) only two out of four tetrahedral bonds of each Si atom are excited^{17,20}. This real-space anisotropy translates to the excitation of electrons to four out of eight equivalent local minima (L-valleys) and holes along Λ lines in reciprocal space. This anisotropic distribution exerts a force on the lattice through an imbalance of deformation potentials, which describe distortion of the lattice in response to charge density fluctuations^{17,18}, and of chemical potentials between different nonthermal carrier populations²⁰. These two carrier–lattice interactions, which contribute to the imaginary part of the Raman tensor, exert a step function electrostrictive force on the lattice driving the coherent LO phonon oscillation polarized in the [001] direction¹⁸. By contrast, equivalent excitation of each bond in the Γ_{12} geometry balances the forces to zero, precluding the LO phonon excitation (Fig. 1d inset).

In addition to absorption, electronic Raman scattering contributes microscopic currents to the longitudinal polarization^{11,16,20–22}. The anisotropic inelastic intraband scattering of light by free electrons, which is unscreened, occurs in materials with non-spherical Fermi surfaces—for example, the valence and conduction bands of Si (refs 23, 24). In addition to intraband scattering, the inter-valence band electronic Raman scattering among the heavy-hole, light-hole and split-off bands indicated in Fig. 2c also contribute significantly to the $\chi^{(2)}$ response predominantly in the $\Gamma_{25'}$ geometry²⁵. These electronic contributions have distinct spectral signatures for thermal carriers in n- and p-type Si^{16,21,24}. Intraband scattering measures the carrier velocity distribution with the energy transfer given by $\Delta\omega \approx qV$, where q is the scattering wavevector and V the carrier velocity^{23,24}. The intraband spectrum forms a lorentzian or gaussian wing on the elastic scattering peak with a carrier density and temperature-dependent width of typically 6 THz (ref. 21). When calculated with realistic band structure of Si, the interband component has a single broad maximum at ~20 THz (refs 25, 26). Because it overlaps with the intraband component, the interband component can be identified clearly only for hole doping of $>10^{20} \text{ cm}^{-3}$ (ref. 22). Because the photogenerated carrier density is significantly higher than the doping density, the electronic Raman contributions are mainly from the photogenerated non-thermal plasma. Therefore, electronic spectral distributions are modified from the thermal ones, and their contribution rises with the photogenerated plasma density.

To identify different contributions to the coherent response of Si that appear in the CWT chronograms, Fig. 3 presents cross-sections

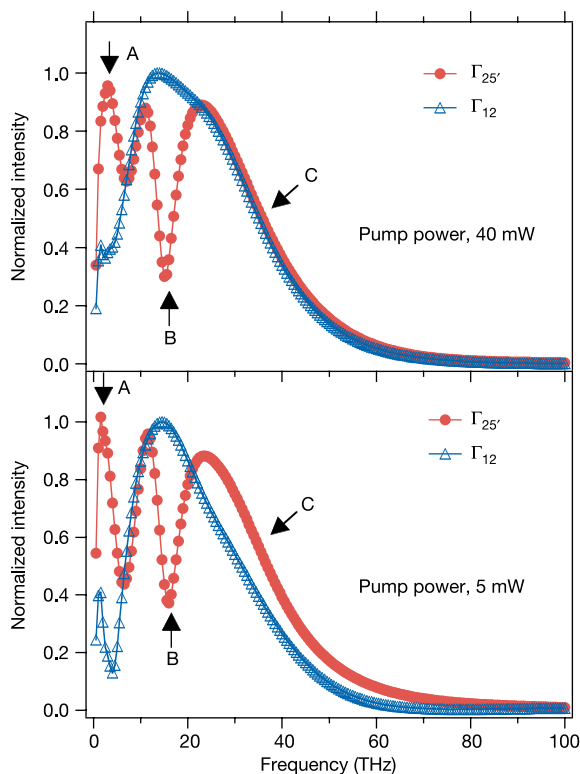


Figure 3 Slices of continuous wavelet transforms (CWTs) in Fig. 1 at zero delay. Data are shown for pump powers of 40 mW (top panel) and 5 mW (bottom panel). The sections show different components A, B and C (explained in the text) that contribute to the nonlinear polarization. The amplitude suppression at zero frequency is an artefact of the CWT procedure and finite measurement time.

of the $\Gamma_{25'}$ and Γ_{12} chronograms at zero delay for 5- and 40-mW excitation. Three distinct features, labelled A, B and C, can be identified. The component A is assigned to the intraband electronic scattering on the basis of its expected frequency. The calculated interband Raman spectra are qualitatively similar to the main component of the electronic response with the maximum at ~ 13 THz (C in Fig. 3)²⁵. The width of this component broadens for high pump-power, particularly in the Γ_{12} geometry, indicating that it depends on the carrier density and distribution. However, the interband scattering cannot make exclusive contribution to C, because the response is nearly identical in the Γ_{12} and $\Gamma_{25'}$ geometries, whereas the interband scattering is mainly observable in the $\Gamma_{25'}$ geometry. We attribute the C component to all electronic processes that contribute to the $\chi^{(3)}$ response, including electron-hole pair generation (photoabsorption), and electronic Raman scattering from the hot plasma.

However, the most intriguing aspect of the coherent response of Si is the anti-resonance (dip) at 15.3 THz in the $\Gamma_{25'}$ geometry (B in Fig. 3). The anti-resonance in CWT chronograms has maximum negative amplitude in the region of overlap between the coherent phonon and electronic responses at 22 fs and 15.3 THz, suggesting that it is related to the coherent excitation of the coupled phonon-carrier systems. The precise location of the anti-resonance in time and frequency depends on the doping of the sample and the pump power, indicating that it is sensitive to the type of dopant and the distribution of nonthermal carriers. The evolution of the anti-resonance (such as its delayed appearance, width that narrows with delay, and position that shifts to higher frequency) are suggestive of quantum mechanical interference phenomena that arise from energy-time uncertainty, parallel excitation pathways, and finite phase relaxation time of the induced polarization (~ 16 fs from the lorentzian linewidth of 80 meV at 300 K for the E_1 critical point)^{8,20,27}.

We attribute the anti-resonance in the CWT chronograms of Si to the coupling of the LO phonon and electron-hole pair continuum amplitudes via the $\chi^{(3)}$ scattering process and electron-phonon interaction. Such discrete-continuum coupling is generic to many physical systems, and is manifest in the Fano spectral line shapes²⁸. The anisotropic reflectivity measurement in Fig. 1a captures the many-body interactions involving the valance band electron-hole pair and LO phonon excitations that gives rise to the Fano line shapes in the LO phonon Raman spectra of highly p-doped Si^{16,22}. By projecting the data into frequency-time space, we can follow the interactions that give birth to coherent phonons^{16,26,28,29}. The $\Gamma_{25'}$ chronograms represent the joint electronic and phonon scattering probabilities coherently coupled via the second-order Raman transition moment and the electron-phonon interaction. Green's function derivations of discrete-continuum spectral functions attribute anti-resonances to final-state interactions and asymmetric line shapes to interference effects, both stemming from the fundamental interaction that allows phonons to decay into, and be generated from, the electron-hole pair continuum^{26,28,29}. The anti-resonance is a manifestation of final-state interactions through which the LO phonon amplitude is dissolved in the electron-hole pair continuum and the electron-hole pair continuum amplitude is suppressed at the resonance frequency. The asymmetry of line shapes arises from the interference between spectral amplitudes, which has the origin in the $-\arctan$ phase relationship between the electronic and lattice responses²⁸. The electronic force leads the LO phonon response by $\pi/2$ at resonance; on either side of the resonance constructive and destructive interference occurs for $<\pi/2$ and $>\pi/2$, respectively. The distinct coherent LO phonon amplitude emerges only after ~ 100 fs delay when the electronic coherences have substantially dephased. The residual manifestation of electron-phonon interaction is the carrier-density-dependent LO phonon frequency and decay rate representing the real and imaginary parts of self-energy.

The CWT chronograms in Fig. 1 provide a frequency-time survey

of the discrete-continuum interaction, a pervasive feature in physics. The transient reflectivity measurements demonstrate the possibility of observing in real time the quantum mechanical manifestations of carrier-phonon interactions in Si, which until now could only have been deduced from transport measurements and spectral lineshape analysis. The observation of the coherent response to impressed external fields at >50 -THz frequencies defines the fundamental response of silicon, and it opens the way to application and manipulation of internal fields and forces that could lead to applications in ultra-broadband signal processing at frequencies that are intrinsically orders of magnitude faster than conventional electronics. $\square$

Received 6 June; accepted 8 September 2003; doi:10.1038/nature02044.

1. Pines, D. & Nozières, P. *Theory of Quantum Liquids* (Benjamin, New York, 1966).
2. Huber, R. *et al.* How many-particle interactions develop after ultrafast excitation of an electron-hole plasma. *Nature* **414**, 286–289 (2001).
3. Fischer, B. & Hofmann, K. R. A full-band Monte Carlo model for the temperature dependence of electron and hole transport in silicon. *Appl. Phys. Lett.* **76**, 583–585 (2000).
4. Downer, M. C. & Shank, C. V. Ultrafast heating of silicon on sapphire by femtosecond optical pulses. *Phys. Rev. Lett.* **56**, 761–764 (1986).
5. Sjodin, T., Petek, H. & Dai, H.-L. Ultrafast carrier dynamics in silicon: A two-color transient reflection grating study on a (111) surface. *Phys. Rev. Lett.* **81**, 5664–5667 (1998).
6. Sabbah, A. J. & Riffe, D. M. Femtosecond pump-probe reflectivity study of silicon carrier dynamics. *Phys. Rev. B* **66**, 165217 (2002).
7. Buhleier, R., Lüpke, G., Marowsky, G., Gogolak, Z. & Kuhl, J. Anisotropic interference of degenerate four-wave mixing in crystalline silicon. *Phys. Rev. B* **50**, 2425–2431 (1994).
8. Bigot, J. Y., Portella, M. T., Schoenlein, R. W., Cunningham, J. E. & Shank, C. V. Two-dimensional carrier-carrier screening in a quantum well. *Phys. Rev. Lett.* **67**, 636–639 (1991).
9. Goldman, J. R. & Prybyla, J. A. Ultrafast dynamics of laser-excited electron distribution in silicon. *Phys. Rev. Lett.* **72**, 1364–1367 (1994).
10. Pfeifer, T., Dekorsky, T., Kütt, W. & Kurz, H. Generation mechanism for coherent LO phonons in surface-space-charge fields of III-V-compounds. *Appl. Phys. A* **55**, 482–488 (1992).
11. Saeta, P. N., Greene, B. I. & Chuang, S. L. Short terahertz pulses from semiconductor surfaces: The importance of bulk difference-frequency mixing. *Appl. Phys. Lett.* **63**, 3482–3484 (1993).
12. Johnston, M. B., Whittaker, D. M., Corchia, A., Davies, A. G. & Linfield, E. H. Simulation of terahertz generation at semiconductor surfaces. *Phys. Rev. B* **65**, 165301 (2002).
13. Khurgin, J. B. Optical rectification and terahertz emission in semiconductors excited above the band gap. *J. Opt. Soc. Am. B* **11**, 2492–2501 (1994).
14. Sipe, J. E., Mizrahi, V. & Stegeman, G. I. Fundamental difficulty in the use of second-harmonic generation as a strictly surface probe. *Phys. Rev. B* **35**, 9091–9094 (1987).
15. Caumes, J.-P., Videau, L., Rouyer, C. & Freysz, E. Kerr-like nonlinearity induced via terahertz generation and the electro-optical effect in zinc blende crystals. *Phys. Rev. Lett.* **89**, 047401 (2002).
16. Cerdeira, F., Fjeldly, T. A. & Cardona, M. Effect of free carriers on zone-center vibrational modes in heavily doped p-type Si II. Optical modes. *Phys. Rev. B* **8**, 4734–4745 (1973).
17. Scholz, R., Pfeifer, T. & Kurz, H. Density-matrix theory of coherent phonon oscillations in germanium. *Phys. Rev. B* **47**, 16229–16236 (1993).
18. Stevens, T. E., Kuhl, J. & Merlin, R. Coherent phonon generation and the two stimulated Raman tensors. *Phys. Rev. B* **65**, 144304 (2002).
19. Lautenschlager, P., Garriga, M., Viña, L. & Cardona, M. Temperature dependence of the dielectric function and interband critical points in silicon. *Phys. Rev. B* **36**, 4821–4830 (1987).
20. Cerdeira, F. & Cardona, M. Effect of carrier concentration on the Raman frequencies of Si and Ge. *Phys. Rev. B* **5**, 1440–1454 (1972).
21. Contreras, G., Sood, A. K. & Cardona, M. Raman scattering by intervalley carrier-density fluctuations in n-type Si: Intervalley and intravalley mechanisms. *Phys. Rev. B* **32**, 924–929 (1985).
22. Chandrasekhar, M., Rössler, U. & Cardona, M. Intra- and interband Raman scattering by free carriers in heavily doped p-Si. *Phys. Rev. B* **22**, 761–770 (1980).
23. Wolff, P. A. Effect of nonparabolicity on light scattering from plasmas in solids. *Phys. Rev.* **171**, 436–444 (1968).
24. Bairamov, B. H., Ipatova, I. P. & Voitenko, V. A. Raman scattering from current carriers in solids. *Phys. Rep.* **229**, 221–290 (1993).
25. Kanehisa, M. A., Wallis, R. F. & Balkanski, M. Interband electronic Raman scattering in p-silicon. *Phys. Rev. B* **25**, 7619–7625 (1982).
26. Belitsky, V. I., Cantarero, A., Cardona, M., Trallero-Giner, C. & Pavlov, S. T. Feynman diagrams and Fano interference in light scattering from doped semiconductors. *J. Phys. Condens. Matter* **9**, 5965–5976 (1997).
27. Lautenschlager, P., Allen, P. B. & Cardona, M. Phonon-induced lifetime broadenings of electronic states and critical points in Si and Ge. *Phys. Rev. B* **33**, 5501–5511 (1986).
28. Fano, U. Effects of configuration interaction on intensities and phase shifts. *Phys. Rev.* **124**, 1866–1878 (1961).
29. Balkanski, M., Jain, K. P., Beserman, R. & Jouanne, M. Theory of interference distortion of Raman scattering line shapes in semiconductors. *Phys. Rev. B* **12**, 4328–4337 (1975).

Acknowledgements We thank D. Boyanovsky, A. P. Heberle, K. Ishioka and J. Shan for discussions. This work was supported by the NSF, the University of Pittsburgh, a Grant-in-Aid for Scientific Research from MEXT of Japan, and NIMS Research Funds.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.P. (petek@pitt.edu).

Magnetic control of ferroelectric polarization

T. Kimura^{1*}, T. Goto¹, H. Shintani¹, K. Ishizaka¹, T. Arima² & Y. Tokura¹

¹Department of Applied Physics, University of Tokyo, Tokyo 113-8656, Japan

²Institute of Materials Science, University of Tsukuba, Tsukuba 305-8573, Japan

* Present address: Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

The magnetoelectric effect—the induction of magnetization by means of an electric field and induction of polarization by means of a magnetic field—was first presumed to exist by Pierre Curie¹, and subsequently attracted a great deal of interest in the 1960s and 1970s (refs 2–4). More recently, related studies on magnetic ferroelectrics^{5–14} have signalled a revival of interest in this phenomenon. From a technological point of view, the mutual control of electric and magnetic properties is an attractive possibility¹⁵, but the number of candidate materials is limited and the effects are typically too small to be useful in applications. Here we report the discovery of ferroelectricity in a perovskite manganite, TbMnO₃, where the effect of spin frustration causes

sinusoidal antiferromagnetic ordering. The modulated magnetic structure is accompanied by a magnetoelastically induced lattice modulation, and with the emergence of a spontaneous polarization. In the magnetic ferroelectric TbMnO₃, we found gigantic magnetoelectric and magnetocapacitance effects, which can be attributed to switching of the electric polarization induced by magnetic fields. Frustrated spin systems therefore provide a new area to search for magnetoelectric media.

The room-temperature crystal structure of TbMnO₃ investigated here is the orthorhombically distorted perovskite structure (space group *Pbnm*; Fig. 1a). We note that the perovskite structure of TbMnO₃ is distinct from that of the antiferromagnetic (AF) ferroelectric hexagonal rare-earth manganites (for example, YMnO₃) in which magnetoelectric coupling has been well studied^{5–9}. Furthermore, the perovskite TbMnO₃ does not contain 6s lone pairs, which produce a polar structure in magnetic ferroelectric perovskites BiMnO₃ (refs 11–13) and BiFeO₃ (ref. 14). The electronic configuration of the Mn³⁺ site in TbMnO₃ is identical with that in the parent compound of colossal magnetoresistive manganites, LaMnO₃, having the $t_{2g}^3e_g^1$ configuration. In LaMnO₃, staggered orbital order of the $d_{3x^2-y^2}/d_{3y^2-r^2}$ type is responsible for the layered (A-type) AF order. In contrast, the spin structure in TbMnO₃ is a sinusoidal AF ordering of the Mn³⁺ moments that takes place at

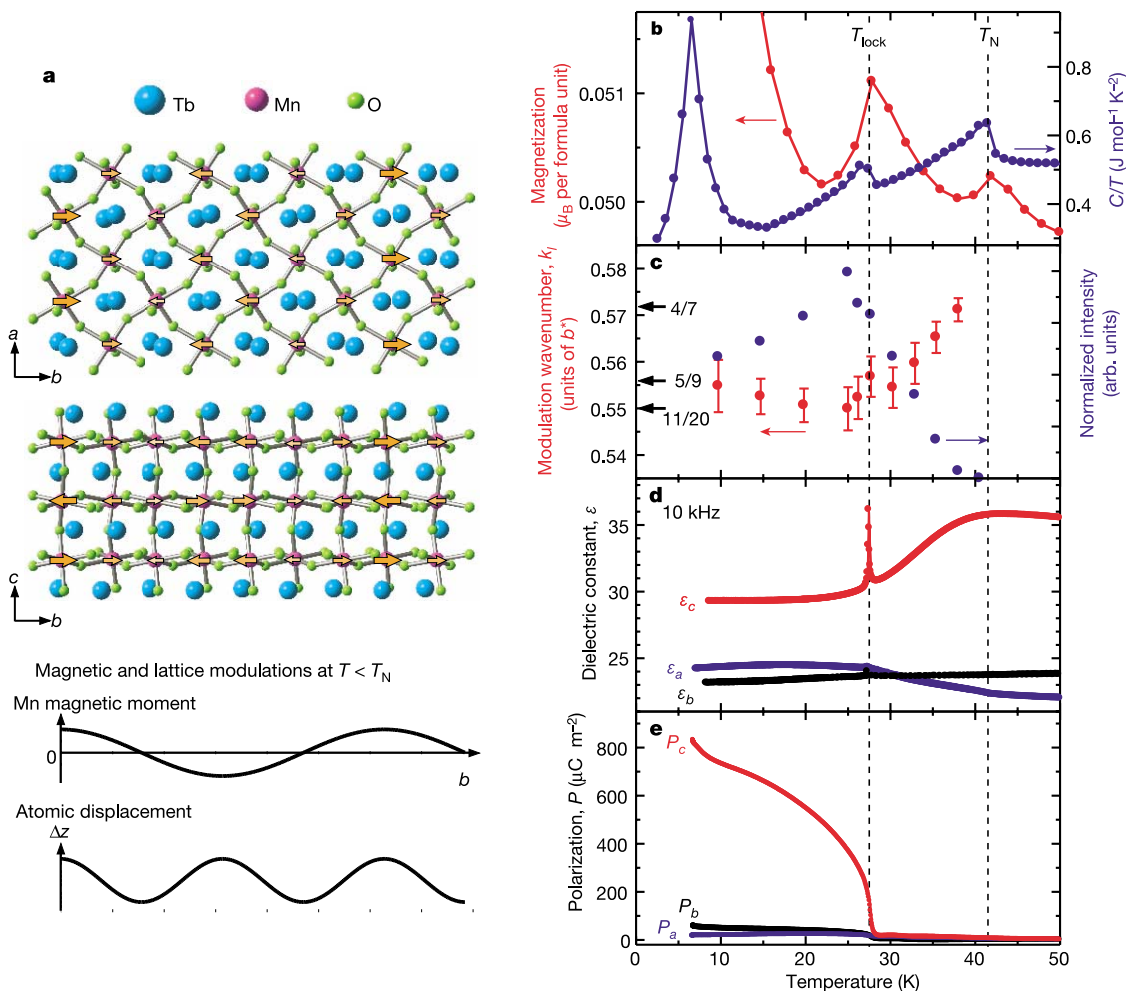


Figure 1 Appearance of ferroelectricity below a lock-in transition temperature in TbMnO₃. **a**, Rough sketches of crystal structure at room temperature (top) and spatial variation along the *b* axis of Mn magnetic moment and atomic displacement ($\Delta z/c$) below T_N (lower). Orange arrows denote Mn magnetic moments below T_N . **b–e**, Temperature

profiles of magnetization and specific heat divided by temperature C/T (**b**), wavenumber of lattice modulation k_1 (**c**), dielectric constant ϵ at 10 kHz (**d**), and electric polarization P along the principal axes in single crystals of TbMnO₃ (**e**). The error bars of k_1 denote the distribution of k related to the correlation length.

$T_N \approx 41$ K with a wave vector $\mathbf{q} = (0, k_s, 1)$ in the $Pbnm$ orthorhombic cell¹⁶. A Mn^{3+} moment oriented along the b axis has been proposed¹⁶. The wavenumber k_s is incommensurate (~ 0.295) at T_N and decreases with decreasing temperature, becoming nearly constant (0.28) below ~ 30 K. A rough sketch of the magnetic order on Mn moments is illustrated in Fig. 1a. Recently, the emergence of long-period sinusoidal AF order in $TbMnO_3$ has been explained in terms of spin frustration caused by the combination of a $GdFeO_3$ -type distortion and staggered orbital order¹⁷.

We show in Fig. 1b temperature (T) profiles of magnetization (M) at 0.5 T, and of specific heat divided by temperature (C/T), for a single crystal of $TbMnO_3$. Both M and C/T exhibit three anomalies at ~ 7 K, ~ 27 K and ~ 42 K. Taking account of results of a former neutron diffraction study¹⁶, the anomalies at ~ 7 K and ~ 42 K are attributed to the magnetic ordering of the Tb^{3+} moments and the sine-wave ordering of the Mn^{3+} moments, respectively. The anomaly at 27 K seems to be related to the incommensurate-commensurate (or lock-in) transition where the magnetic modulation wave vector k_b is locked at a constant value.

X-ray diffraction measurements revealed that the modulated magnetic ordered phase is accompanied by magnetoelastically induced lattice modulation¹⁷. Figure 1c displays the T dependence of wavenumber k_l and normalized intensity of a superlattice reflection at $(0, k_l, 3)$ attributed to the lattice modulation. The superlattice reflections at the wave vector $(0, k_l, l)$ with l an integer, appear below $T_N \approx 41$ K ($k_l \approx 0.57$ at T_N). The intensity of the superlattice reflections increases down to the lock-in transition temperature ($T_{lock} \approx 27$ K), but is suppressed again below T_{lock} . Another notable feature is the T dependence of the modulation wavenumber k_b , as clarified in Fig. 1c. Upon cooling from T_N down to T_{lock} , k_b decreases and then becomes nearly constant ($k_b \approx 0.55$) below T_{lock} . The value of the T -dependent k_l is almost twice as large as that of the magnetic wavenumber k_s (ref. 16). It is well known that the crystallographic deformations upon magnetic ordering are caused by the magnetic atoms increasing their exchange interaction energy by shifting their positions; that is, exchange striction^{18,19}. By analogy with the lattice modulations observed in some rare-earth metals such as Tm with sinusoidal order^{20,21}, the observed super-

lattice reflections due to atomic displacement can be regarded as a second harmonic that is magnetoelastically induced by sinusoidal AF order.

In materials with incommensurate phases²², such as the family of A_2BX_4 compounds (for example, Rb_2ZnCl_4 , $(NH_4)_2BeF_4$ and K_2SeO_4), the lattice modulation is generally connected with a spatially varying electric polarization. With further decreasing temperature, these materials exhibit a first-order phase transition (lock-in transition) to a ferroelectric phase with finite spontaneous polarization. Taking account of these observations, we could expect to find ferroelectricity in $TbMnO_3$ associated with the incommensurate phase at the lock-in transition. Thus, we performed measurements of the dielectric constant, ϵ , and the electric polarization, P , along the a , b and c axes. Figure 1d displays T profiles of ϵ at 10 kHz. While ϵ_b (electric field E parallel to the b axis; $E||b$) shows merely weak T dependence, several anomalies are observed in ϵ_a ($E||a$) and ϵ_c ($E||c$). ϵ_a increases below $T_N \approx 41$ K towards low temperature, and becomes nearly constant below T_{lock} . The most striking feature is observed in ϵ_c . A pronounced λ -type peak of ϵ_c is found at T_{lock} , although no distinct anomaly is observed at T_N . The sharp peak structure in ϵ_c at T_{lock} is suggestive of the occurrence of the ferroelectric phase transition. To confirm the ferroelectric activity, we show in Fig. 1e the temperature variation of P . A finite spontaneous polarization appears below T_{lock} along the c axis. P_c ($P||c$) at ~ 10 K is about 8×10^{-4} C m⁻². We also confirmed that P_c can be reversed by the d.c. electric field. Thus, $TbMnO_3$ becomes ferroelectric below T_{lock} .

The spontaneous polarization of $TbMnO_3$ is rather small compared with that of conventional ferroelectrics (for example, $\sim 2.6 \times 10^{-2}$ C m⁻² at 296 K in $BaTiO_3$), but is comparable to that of the so-called improper ferroelectrics²³ (for example, $\sim 1.2 \times 10^{-3}$ C m⁻² at 153 K in Rb_2ZnCl_4 (ref. 24) and $\sim 5.6 \times 10^{-4}$ C m⁻² at 77 K in K_2SeO_4 (ref. 25)). In improper ferroelectrics, the primary order parameter represents the lattice distortion mode having a nonzero wave vector (that is $k_l \neq 0$), and the spontaneous polarization appears as a secondary order parameter induced by the lattice distortion. This may also be the case for $TbMnO_3$. An uncompensated antipole structure is possible in

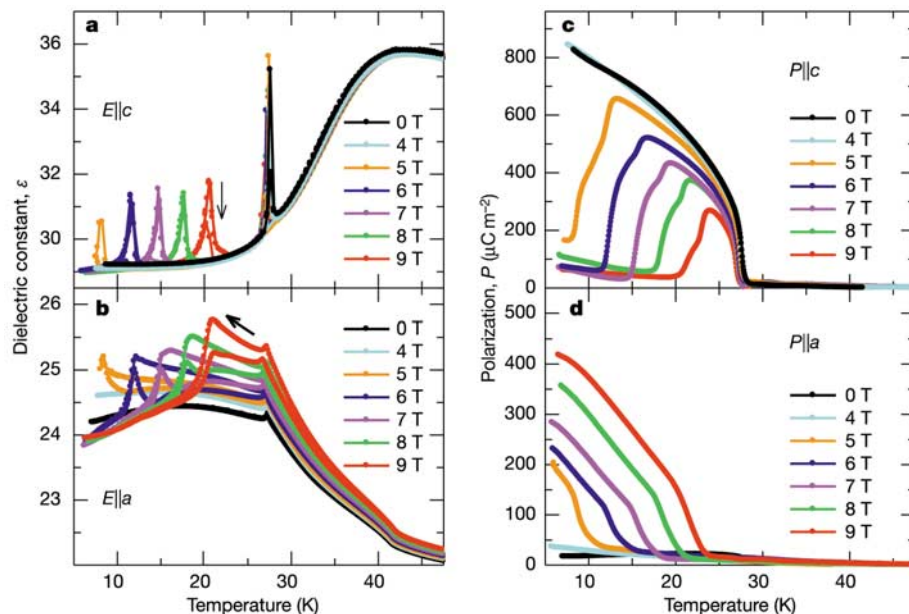


Figure 2 Electric polarization flop induced by magnetic fields in $TbMnO_3$. **a–d**, Temperature profiles of dielectric constant at 10 kHz (**a** and **b**) and of electric

polarization along the c and a axes (**c** and **d**), respectively, at various magnetic fields in single crystals of $TbMnO_3$. Magnetic fields are applied along the b axis.

improper ferroelectrics, which was specified by the term ferrielectricity²⁶ because of its analogy with ferrimagnetism. Indeed, the intensity comparison of the superlattice reflections at $(0, k_l, 0)$ and $(0, k_l, 2)$ measured by the X-ray diffraction suggests that the atomic displacement in the ferroelectric phase has a component along the c axis, that is, the direction of the spontaneous polarization (Fig. 1a). In addition, weak intensities of the superlattice reflections (five orders of magnitude smaller than the Bragg peak) and the lack of their enhancement on tuning the X-ray photon energy at Mn K - and Tb L -absorption edges suggest the lattice distortion is due to the displacement of O ions.

As the lattice modulation in TbMnO₃ is accompanied by magnetic order, we may expect coupling between magnetization and polarization. We have investigated the effect of magnetic field on ϵ and P . In the experiments, magnetic fields (B) were applied along the b axis, that is, the direction of the modulation wave vector and the Mn³⁺ magnetic moments. Figure 2a and b shows the T profiles of ϵ along the c and a axes, respectively, at various magnetic fields. No distinct change of ϵ for either direction was observed at temperatures above T_{lock} . Although the sharp peak at T_{lock} in ϵ_c is not sensitive to the application of magnetic fields, another peak feature at $T_{\text{flop}}(B)$ shows up in ϵ_c below T_{lock} above 5 T. The magnetic-field-induced sharp maximum is shifted towards higher T with

increasing magnetic field from $T_{\text{flop}}(5 \text{ T}) \approx 8 \text{ K}$ to $T_{\text{flop}}(9 \text{ T}) \approx 20 \text{ K}$. The component ϵ_a also shows a remarkable magnetic field effect. By applying magnetic fields above 5 T, ϵ_a exhibits a peak structure at T_{flop} , and shows a thermal hysteresis at a temperature between T_{flop} and T_{lock} . We also measured the ϵ_b in magnetic fields, but no substantial magnetic field effect was observed up to 9 T.

Measurements of P in magnetic fields have revealed the origin of the magnetic-field-induced anomaly in ϵ . We show in Fig. 2c and d the T variation of P_c and P_a , respectively, at various magnetic fields. By applying a magnetic field above $\sim 5 \text{ T}$, the spontaneous polarization in P_c is considerably suppressed below T_{flop} . With increasing magnetic field, the temperature region with finite P_c shrinks, corresponding to the shift of T_{flop} . By contrast, the application of a magnetic field induces a finite P_a (Fig. 2d). The onset temperature of a finite P_a coincides well with T_{flop} , and increases in accord with the increase of T_{flop} by applying a magnetic field. These results show that the spontaneous polarization is switched ('flopped') from the direction along the c axis to the direction along the a axis at T_{flop} , and the strong magnetic field dependence of T_{flop} causes the notable magnetic field effect on ϵ and P .

We display in Fig. 3 the magnetic field dependence of the change in ϵ ($\Delta\epsilon(B)/\epsilon(0) = [\epsilon(B) - \epsilon(0)]/\epsilon(0)$), P and magnetization at selected temperatures. As shown in Fig. 3a and b, a remarkable

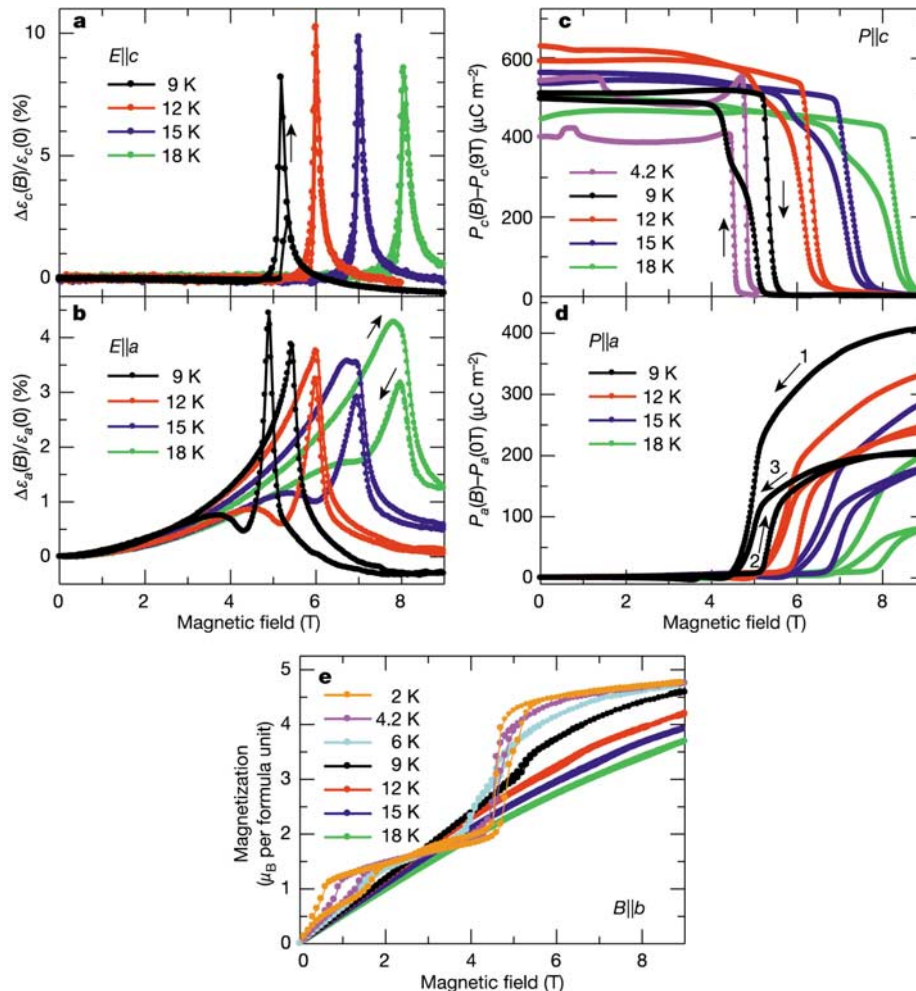


Figure 3 Magnetocapacitance and magnetoelectric effects in TbMnO₃. **a–e**, Magnetic-field-induced change in dielectric constant (**a** and **b**), electric polarization along the c and a axes, respectively (**c** and **d**), and magnetization at selected temperatures (**e**). Magnetic fields are applied along the b axis. The magnetic field dependence of electric

polarization was obtained by the measurements of magnetoelectric current, which were performed after poling crystals. The data of **d** were collected after the magnetic field cooling. The numbers in **d** denote the order of measurements at 9 K.

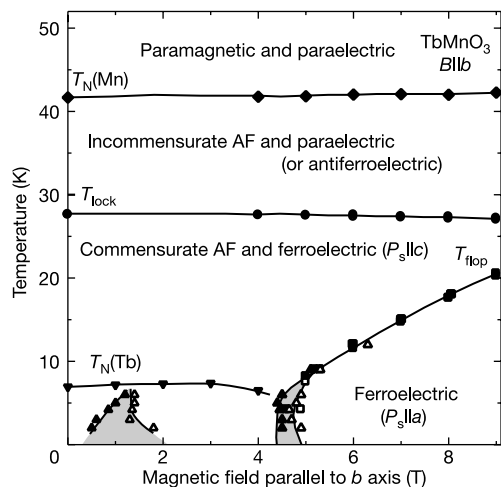


Figure 4 Temperature versus magnetic field phase diagram for TbMnO_3 for magnetic field applied along the b axis. Open and filled symbols represent the data points in the cooling (or magnetic-field increasing) and warming (or magnetic-field decreasing) runs, respectively. $T_N(\text{Mn})$ (determined from the dielectric anomaly in ϵ_a) and $T_N(\text{Tb})$ (determined from the anomaly in the M - T curve) indicate the antiferromagnetic ordering temperature of the Mn^{3+} and Tb^{3+} moments, respectively. T_{lock} and T_{flop} , which were determined from the dielectric anomaly, denote the temperatures of incommensurate–commensurate (or lock-in) transition and electric polarization flop, respectively. Triangles indicate the points where the magnetization curves show steep steps. The shaded areas show magnetic field hysteresis regions.

peak structure with hysteresis is observed at almost the same magnetic field for both $\Delta\epsilon_c(B)/\epsilon_c(0)$ and $\Delta\epsilon_a(B)/\epsilon_a(0)$ at their respective temperatures. The peak position $B_{\text{flop}}(T)$ is shifted towards higher magnetic field with increasing temperature. The maximum value of the magnetocapacitance as defined by $\Delta\epsilon(B_{\text{flop}})/\epsilon(0)$ reaches values as large as $\sim 10\%$ at 12 K. The divergent increase of ϵ at B_{flop} has an intimate connection to the magnetic-field-induced electric polarization flop, which is confirmed by the magnetic field dependence of P (Fig. 3c and d). The spontaneous polarization along the c axis is suddenly suppressed at B_{flop} , whereas that along the a axis appears at B_{flop} with increasing magnetic fields. The data in Fig. 3c and d can be regarded as a novel magnetoelectric effect, where the gigantic change in P ($\Delta P(B) \approx 6 \times 10^{-4} \text{ C m}^{-2}$) is caused by the magnetic-field-induced electric polarization flop with the nature of first-order phase transition.

To sum up the present study, we have determined the electric and magnetic phase diagram of TbMnO_3 from dielectric and magnetic anomalies for the direction of magnetic field parallel to the b axis (Fig. 4). The magnetic phase boundaries denoted by triangles are determined from step steps in magnetization curves (Fig. 3e). This clearly indicates that $T_{\text{flop}}(B)$ is in good agreement with a magnetic phase boundary. Similar metamagnetic transitions also take place in a rare-earth orthoferrite, TbFeO_3 , and this has been interpreted as the magnetic reversal of Ising Tb^{3+} moments^{27,28}. Furthermore, the Tb^{3+} spin reversal gives rise to the spin reorientation in Fe sites. By analogy, it is likely in the magnetic-field-induced electric polarization flop of TbMnO_3 that Tb^{3+} moment reversal, accompanied by the Mn spin reorientation, changes the exchange interaction energy and then brings about the lattice modulation owing to a finite spontaneous polarization along the a axis. The giant magnetocapacitance and magnetoelectric effects as observed in the present frustrated-spin system may provide intriguing opportunities to explore a new type of magnetoelectric functionality. □

Methods

Single crystals of TbMnO_3 were grown in a flow of Ar by the floating zone method. The crystals were oriented using Laue X-ray diffraction patterns, and cut into thin plates (typical dimensions, $\sim 3 \times 3 \times 1 \text{ mm}^3$) with the widest faces perpendicular to the crystallographic principal axes. The magnetization and specific heat for the crystals were measured with a commercial magnetometer and a relaxation technique, respectively. Single-crystal diffraction measurements were performed at BL-4C of the Photon Factory, KEK, in Tsukuba. X-rays with photon energy of 13 keV were used for the non-resonant X-ray scattering, and that near the Mn K - and Tb L -absorption edges for the resonant one. The dielectric constant was measured at 10 kHz using an LCR meter. The T dependence of electric polarization was obtained by measurements of pyroelectric current. The sample was cooled down while applying a poling field ($\sim 150 \text{ kV m}^{-1}$). At the lowest T , the poling electric field was removed. Then, the sample was heated at a constant rate (5 K min^{-1}), and the pyroelectric current was measured. For these electric measurements, silver electrodes were evaporated onto the widest faces of the sample.

Received 26 May; accepted 26 August 2003; doi:10.1038/nature02018.

- Curie, P. Sur la symétrie dans les phénomènes physiques, symétrie d'un champ électrique et d'un champ magnétique. *J. Phys.* 3 (Ser. III), 393–415 (1894).
- O'Dell, T. H. *The Electrodynamics of Magneto-Electric Media* (North-Holland, Amsterdam, 1970).
- Freeman, A. J. & Schmid, H. (eds) *Magnetoelectric Interaction Phenomena in Crystals* (Gordon and Breach, London, 1975).
- Smolenskii, G. A. & Chupis, I. E. Ferroelectromagnets. *Usp. Fiz. Nauk.* 137, 415–448 (1982); also *Sov. Phys. Usp.* 25, 475–493 (1982).
- Huang, Z. J., Cao, Y., Sun, Y., Xue, Y. Y. & Chu, C. W. Coupling between the ferroelectric and antiferromagnetic orders in YMnO_3 . *Phys. Rev. B* 56, 2623–2626 (1997).
- Iwata, N. & Kohn, K. Dielectric anomalies at magnetic transitions of hexagonal rare earth manganese oxides RMnO_3 . *J. Phys. Soc. Jpn* 67, 3318–3319 (1998).
- Katsufuji, T. et al. Dielectric and magnetic anomalies and spin frustration in hexagonal RMnO_3 ($R = \text{Y, Yb, and Lu}$). *Phys. Rev. B* 64, 104419 (2001).
- Fiebig, M., Lottermoser, Th., Fröhlich, D., Goltsev, A. V. & Pisarev, R. V. Observation of coupled magnetic and electric domains. *Nature* 419, 818–820 (2002).
- Hanamura, E., Hagita, K. & Tanabe, Y. Clamping of ferroelectric and antiferromagnetic order parameters of YMnO_3 . *J. Phys. Condens. Matter* 14, L103–L109 (2003).
- Zhong, C. G. & Jiang, Q. The study of the coupling mechanism between antiferromagnetic and ferroelectric ordering and thermodynamic properties in ferroelectromagnets. *J. Phys. Condens. Matter* 14, 8605–8612 (2002).
- Seshadri, R. & Hill, N. A. Visualizing the role of Bi 6s “lone pairs” in the off-center distortion in ferromagnetic BiMnO_3 . *Chem. Mater.* 13, 2892–2899 (2001).
- Moreira dos Santos, A. et al. Evidence for the likely occurrence of magnetoferroelectricity in the simple perovskite, BiMnO_3 . *Solid State Commun.* 122, 49–52 (2002).
- Kimura, T. et al. Magnetocapacitance effect in multiferroic BiMnO_3 . *Phys. Rev. B* 67, 180401(R) (2003).
- Wang, J. et al. Epitaxial BiFeO_3 multiferroic thin film heterostructures. *Science* 299, 1719–1722 (2003).
- Wood, V. E. & Austin, A. E. Possible applications for magnetoelectric materials. *Int. J. Magn.* 5, 303–315 (1974).
- Quezel, S., Tcheou, F., Rossat-Mignod, J., Quezel, G. & Roudaut, E. Magnetic structure of the perovskite-like compound TbMnO_3 . *Physica B* 86–88, 916–918 (1977).
- Kimura, T. et al. Distorted perovskite with e_g configuration as a frustrated spin system. *Phys. Rev. B* 68, 060403(R) (2003).
- Greenwald, S. & Smart, J. S. Deformations in the crystal structures of anti-ferromagnetic compounds. *Nature* 166, 523–524 (1950).
- Smart, J. S. & Greenwald, S. Crystal structure transitions in antiferromagnetic compounds at the Curie temperature. *Phys. Rev.* 82, 113–114 (1951).
- Bohr, J. et al. Diffraction studies of rare earth metals and superlattices. *Physica B* 159, 93–105 (1989).
- Bohr, J., Gibbs, D. & Huang, K. X-ray-diffraction studies of the magnetic state of thulium. *Phys. Rev. B* 42, 4322–4328 (1990).
- Blin, R. & Levanyuk, A. P. *Incommensurate Phases in Dielectrics 1. Fundamentals* (North-Holland, Amsterdam, 1986).
- Levanyuk, A. P. & Sannikov, D. G. Improper ferroelectrics. *Usp. Fiz. Nauk.* 112, 561–589 (1974); also *Sov. Phys. Usp.* 17, 199–214 (1974).
- Sawada, S., Shiroishi, Y., Yamamoto, A., Takashige, M. & Matsuo, M. Ferroelectricity in Rb_2ZnCl_4 . *J. Phys. Soc. Jpn* 43, 2099–2100 (1977).
- Aiki, K., Hukuda, K. & Matsumura, O. Ferroelectricity in K_2SeO_4 . *J. Phys. Soc. Jpn* 26, 1064 (1969).
- Goldsmith, G. J. & White, J. G. Ferroelectric behavior in thiourea. *J. Chem. Phys.* 31, 1175–1187 (1959).
- Bouree, J. E. & Hamann, J. Mise en évidence expérimentale des effets de forme dans l'orthoferrite de terbium. *J. Phys.* 36, 391–397 (1975).
- Bidaux, R., Bouree, J. E. & Hamann, J. Diagramme de phase de l'orthoferrite de terbium en présence d'un champ magnétique. *J. Phys.* 36, 803–809 (1975).

Acknowledgements We thank K. Kohn, K. Ohgushi, S. Ishihara and A. P. Ramirez for discussions, and Y. Wakabayashi for help with X-ray diffraction measurements. This work was partly supported by KAKENHI from the MEXT of Japan.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.K. (tkimura@lanl.gov).

Self-assembly in aqueous solution of wheel-shaped Mo₁₅₄ oxide clusters into vesicles

Tianbo Liu¹, Ekkehard Diemann³, Huilin Li², Andreas W. M. Dress⁴ & Achim Müller³

¹Department of Physics, ²Department of Biology, Brookhaven National Laboratory, Upton, New York 11973, USA

³Faculty of Chemistry, ⁴Faculty of Mathematics, University of Bielefeld, D-33501 Bielefeld, Germany

Surfactants and membrane lipids readily assemble into complex structures¹ such as micelles, liposomes or hollow vesicles owing to their amphiphilic character—the fact that part of their structure is attracted to polar environments while another part is attracted to non-polar environments. The self-assembly of complex structures also occurs in polyoxometallate chemistry, as exemplified by the molybdenum blue solutions known for centuries. But while the presence of nanometre-sized metal oxide aggregates in these solutions has long been recognized, unravelling the composition and formation process of these aggregates proved difficult. Recent work has indicated that discrete, wheel-shaped mixed-valence polyoxomolybdate clusters of the type {Mo₁₅₄} (refs 2–4) assemble into well-defined nanometre-sized aggregates, including spherical structures⁵. Here we report light-scattering data and transmission electron microscopy images of hollow spherical structures with an average, almost monodisperse radius of about 45 nm and composed of approximately 1,165 {Mo₁₅₄} wheel-shaped clusters. The clusters appear to lie flat and homogeneously distributed on the vesicle surface. Unlike conventional lipid vesicles, the structures we observe are not stabilized by hydrophobic interactions. Instead, we believe the polyoxomolybdate-based vesicles form owing to a subtle interplay between short-range van der Waals attraction and long-range electrostatic repulsion, with important further stabilization arising from hydrogen bonding involving water molecules encapsulated between the wheel-shaped clusters and in the vesicles' interior.

Centuries ago, Native Americans observed a mysterious “blue water”, which is now thought to have been the molybdenum blue solution formed naturally upon oxidation of molybdenite, MoS₂ (ref. 4). The presence of nanometre-sized metal oxide aggregates in solution can easily be proved by the Tyndall effect and trying to unravel the exact composition and structure of the dissolved entities has long been a challenging problem, attracting the interest of chemists such as Scheele⁶ and Berzelius⁷. But firm insights into the complex molecular structure of the anionic units have emerged only over the last decade, following controlled synthesis and subsequent precipitation of giant wheel-shaped {Mo₁₅₄} anions⁴ derived from molybdenum blue (see Fig. 1a–c). Because these anions have highly hydrophilic surfaces that render them extremely soluble, they are preferably precipitated under high concentration or through a salting-out process^{2–4,8}, analogous to the salting-out of proteins. The giant wheel-shaped anions have a composition almost identical to that of natural molybdenum blue (the mineral ilsemanite, approximate formula Mo₃O₈·nH₂O)⁹. A preliminary study indicated that they aggregate into large spherical assemblies that burst in vacuum⁵, suggesting a hollow, vesicle-like aggregate structure.

To further elucidate the composition and structure of the aggregates, we performed static and dynamic light scattering (SLS and DLS) measurements a few days after their aggregation and growth in water was complete. The parent solution was prepared

using the well-defined crystalline product Na₁₅[Mo(VI)₁₂₆Mo(V)₂₈O₄₆₂H₁₄(H₂O)₇₀]_{0.5}[Mo(VI)₁₂₄Mo(V)₂₈O₄₅₇H₁₄(H₂O)₆₈]_{0.5} · 400 H₂O, which contains, beside Na⁺ and H⁺ ions, {Mo₁₅₄}[–] and {Mo₁₅₂}[–]-type anionic units (the latter unit has properties identical to those of the former, and is missing only one Mo₂ unit)^{2–4}. The results of a CONTIN analysis¹⁰ of DLS measurement performed on the aqueous {Mo₁₅₄} solution are shown in the inset of Fig. 2, demonstrating that the aggregates have an almost monodisperse size, with an average hydrodynamic radius *R*_h of ~45 nm. SLS data, analysed through the Zimm plot¹¹, indicate that the average radius of gyration *R*_g of the aggregates is 45.2 ± 1.4 nm (Fig. 2), that is, the two radii are identical. For a solid spherical particle, the relationship between the radii is *R*_g = 0.77*R*_h. If our aggregates were solid spheres with *R*_h = 45 nm, we would expect an *R*_g value of ~34 nm, which is clearly not the case. As an increasing fraction of the total mass of a spherical object is distributed closer to the sphere's surface, the ratio *R*_g/*R*_h increases; the ratio approaches a value of 1 as the corresponding spherical object has all its mass on its surface, the typical model for a vesicle structure¹². Our scattering data thus indicate that the aggregates in solution are not ‘solid clusters’, but vesicle-like hollow spheres.

The mass of the aggregates (*M*_w) determined from the Zimm plot is ~ (2.54 ± 0.25) × 10⁷ g mol^{–1}, which corresponds to ~1,165 individual nanowheels. Because a solid {Mo₁₅₄}/{Mo₁₅₂}-type nanocrystal with a 45-nm radius would contain more than 14,000 individual molecules, the estimated *M*_w value also suggests that our aggregates must have a hollow interior. As a model (see

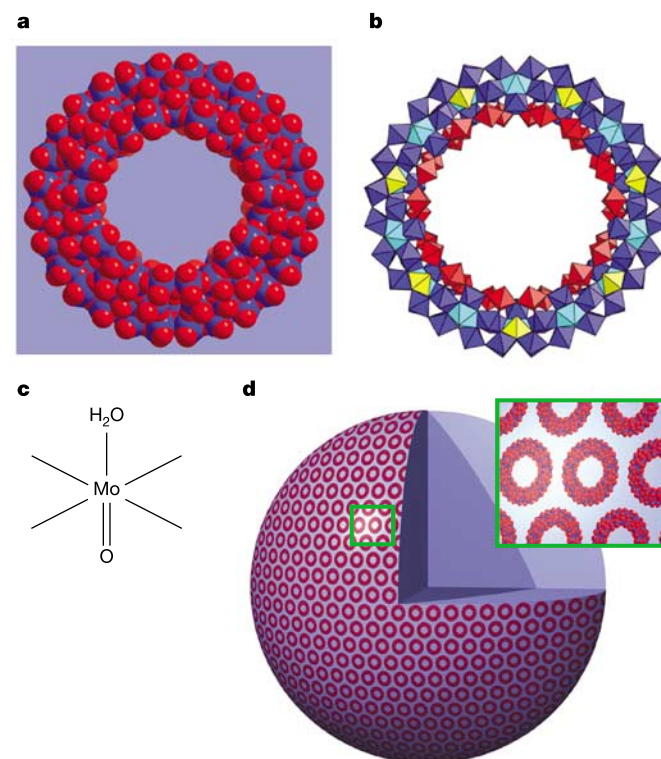


Figure 1 Structure of the 3.6 nm size {Mo₁₅₄}-type nanowheel with a hydrophilic surface and nanosized central cavity. **a**, Space-filling representation (blue and light blue, Mo atoms; red, O atoms). **b**, Polyhedral representation, demonstrating the abundance of pentagonal (Mo)₅ units (in blue) probably influencing the water structure (Mo₂ units in red, Mo₁ units in yellow). **c**, The typical smallest fragment with a metal atom and its coordination sphere, that is, with one of the 70 H₂O ligands causing the extreme hydrophilic nature that is responsible for the interaction with solvents such as water. **d**, Schematic plot of the vesicle structure formed from nanowheels (~45 nm radius) in aqueous solution, with inset showing enlarged nanowheels.

illustration in Fig. 1d), we suggest that $\sim 1,165$ nanowheels are homogeneously distributed on the surface of a vesicle, with their molecular isotropic x - y plane parallel to the surface. If the x - y plane were perpendicular to the vesicle surface, the nanowheels would be expected to exhibit tubular rather than spherical morphology. In fact, a tubular morphology has been reported for the liquid crystal structure obtained from the present nanowheels encapsulated in surfactants¹³.

Assuming that all the nanowheels are distributed on the surface by almost hexagonal closest packing, the centre-to-centre distance between two adjacent nanowheels would be $\sim 4.9 \pm 0.4$ nm, whereas the diameter of an individual nanowheel is ~ 3.6 nm². This suggests that the nanowheels are not touching each other on the vesicle surface, and that water plays an important role in filling gaps and holding the overall assemblies together. Note that all the water molecules incorporated into the vesicle structures do not contribute to the scattered intensity ($dn/dc = 0$); the M_w value we have estimated from the scattering data does thus not include the weight of any water molecules incorporated into the vesicle. We estimate that if the weight of the water molecules were to be included, the actual M_w of the vesicles would be more than ten times larger than quoted above.

Placing our solution on a support and drying it allowed for high-resolution TEM studies, which indicate the presence of uniform, spherical entities of 35–45 nm in radius, corresponding to the abundance of the vesicles (Fig. 3a). The size range is consistent with the DLS and scanning electron microscope (SEM) measurements. The vesicles tend to burst and collapse during the drying process and hence to lose much water, with the result that the metal-oxide components are closer to each other in the burst product. We occasionally found larger vesicles and chose one of these for a more detailed study that yields better images (Fig. 3b). The burst ('post mortem') vesicle exhibits the wrinkle feature and high contrast around the edge typical for an empty sphere, as seen in case of the lipid vesicles; this observation is suggestive of the soft and flexible nature of the 'surface membrane' of our vesicles. When studying a patch of a broken vesicle flattened on the support carbon film (Fig. 3c), we observed locally ordered packing over small areas but no long-range order, as expected (Fig. 3d, e). Power diffraction spectra taken on these samples clearly indicate some ordering with

lattice spacings of 4.2, 4.2 and 3.6 nm, respectively (Fig. 3f, g). Fourier filtering yields the images (Fig. 3h, i) while the real space distance ~ 3.6 nm corresponds exactly to the wheel diameter, which indicates that the wheels are in direct non-covalent contact after the vesicle has burst following loss of water in the high-vacuum condition.

In contrast to surfactant vesicles assembled by hydrophobic interactions, the present vesicles consist of components that do not have any obvious hydrophobic parts and also keep a considerable distance from each other in the aggregate. We believe that the driving force for the assembly of these structures is a delicate balance between short-range attractive forces (van der Waals forces and hydrogen-bonding forces) and repulsive electrostatic interactions between adjacent anion clusters. This view is supported by the fact that vesicle size can be tuned by adjusting the pH of the solution or adding electrolytes, with larger vesicles ($R_h > 100$ nm) observed at lower pH conditions or in the presence of additional electrolyte such as NaCl. Considering that van der Waals interactions are relatively weak, additional stabilizing interactions are probably needed to keep the system together. This might be provided by hydrogen bonding involving (probably partly protonated) water molecules incorporated in the vesicle structures.

When structures are formed through the assembly of identical constituents, those exhibiting the highest symmetry tend to be energetically preferred^{14,15}; thus, if surfaces of the present type are formed, highly symmetrical spherical structures might be expected. In the case of the $\{\text{Mo}_{154}\}$ nanowheels arranged on a sphere, symmetry considerations suggest that the triangular net obtained by connecting the centres of any two adjacent nanowheels by a straight line should be an icosadeltahedron, that is, a polytope with icosahedral symmetry where all faces are equilateral triangles and either six or five triangles are found adjacent to each vertex. Icosadeltahedra were first classified by Goldberg and are parameterized by two integers, h and k , that describe the mutual position of centres of five-fold symmetry within the triangular net relative to each other. Defining the triangulation number¹⁶ by $t = h^2 + hk + k^2$, the icosadeltahedron with parameters h, k has exactly $20t$ triangular faces and $2 + 10t$ vertices; 12 of these vertices are adjacent to five triangles and all other vertices are adjacent to six triangles. Consequently, vesicles can be expected to contain $2 + 10t$ wheels for some h, k values with $2 + 10t$ close to 1,165, the number obtained from our measurements. The condition $1,115 < 2 + 10t < 1,215$ holds only for $\{h, k\} = \{8, 4\}$, and $\{9, 3\}$, suggesting a corresponding number of $1,122 = 2 + 10(8^2 + 8 \times 4 + 4^2)$, or $1,172 = 2 + 10(9^2 + 9 \times 3 + 3^2)$ wheels within a vesicle.

Both SEM⁵ and DLS measurements show that when using solvents with lower polarity, larger spherical aggregates form. For example, in acetone, uniform spherical vesicles with $R_h \approx 130 \pm 5$ nm and $R_g \approx 128.1 \pm 5.6$ nm are observed. Each vesicle contains $\sim 19,000$ single components, corresponding to a ~ 3.6 nm centre-to-centre distance, nearly the same as their actual size in single crystals². This suggests that the nanowheels pack rather closely with each other to form a spherical shell in acetone. Their behaviour is in accord with the expectation that, as would any electrolyte, the anionic nanowheels should carry a lower charge in a weaker polar solvent (with cations expected to be correspondingly abundant at the anions' surface), leading to weaker inter-nanowheel repulsive electrostatic interactions, and thus to a smaller inter-nanowheel distance.

Further study is required to gain full insight into the unusual solution behaviour of the nanowheels and understand the nature of the interactions that drive the assembly in dilute solution against entropic effects. But we expect nevertheless that this unusual system will offer intriguing possibilities for further exploration and modification^{4,13,17–20}, given the structural complexity of the nanowheel building blocks, their extended hydrophilic surface, unusual geo-

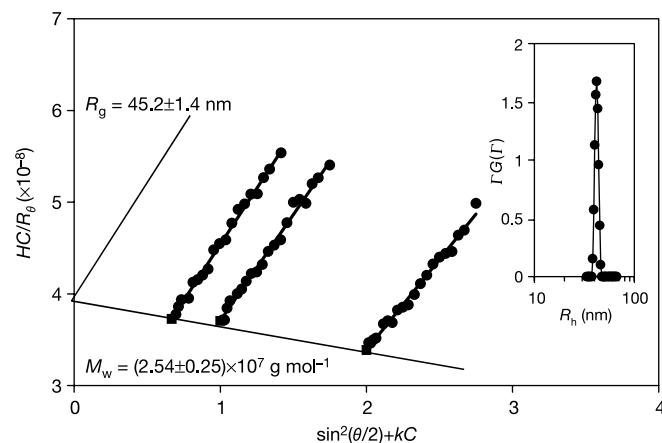


Figure 2 Zimm plot based on SLS measurements. Three different concentrations (from right, 0.03, 0.015 and 0.01 mg ml⁻¹, respectively, at pH = 2.3) are measured. The R_g and the weight-average mass of the aggregates (M_w) can be calculated from the plot (H : an optical parameter; C : vesicle concentration; R_θ : Rayleigh ratio of vesicle solution at scattering angle θ). Inset, CONTIN analysis of DLS study on 0.01 mg ml⁻¹ $\{\text{Mo}_{154}\}/\{\text{Mo}_{152}\}$ aqueous solution at pH = 2.3, scattering angle = 60°. The aggregates in solution have an average R_h of 45 nm, with a very narrow size distribution.

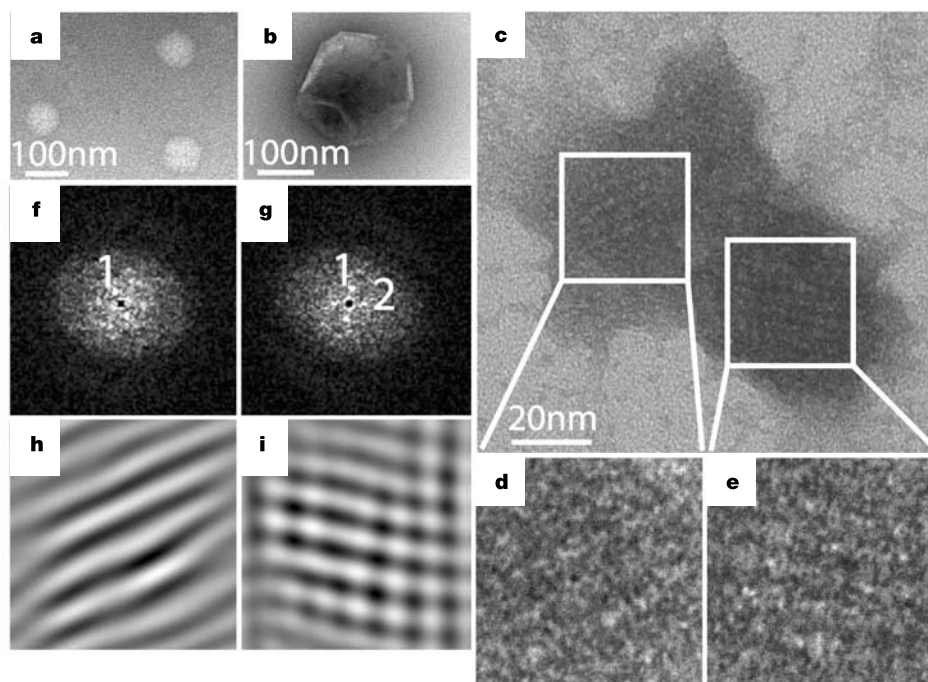


Figure 3 Transmission electron microscopy studies of wheel-type vesicles on carbon film. **a**, 35–45 nm radius vesicles; **b**, image of a larger broken vesicle; **c**, image of a small patch of the broken vesicle flattened on the support carbon film. **d–i**, Sub-areas, demarcated by two square boxes, are zoomed into (**d**, **e**). **f**, **g**, The corresponding Fourier

transforms (power spectra) which clearly show some ordered packing in small local areas (one in **f** and two in **g**). **h**, **i**, Their back-transforms after filtering the high frequencies, showing the ordering of the nanowheels (diameter, 3.6 nm).

metrical structure and large number of delocalized electrons (causing the blue colour)⁴.

Note added in proof: Until immediately before publication, we failed to cite two earlier papers of ours^{21,22} that reported the vesicle formation behaviour of another polyoxometallate system. □

Methods

Static light scattering

A commercial Brookhaven Instrument LLS spectrometer equipped with a solid-state laser operating at 532 nm was used for measurements of both SLS and DLS. SLS experiments were performed at scattering angles between 10 and 120°, at 2° intervals. The basis of the SLS data analysis is the Rayleigh–Gans–Debye equation¹¹ to obtain the radius of gyration (R_g) and the weight-average molar mass (M_w) of the particles. As the present solutions indicate a certain absorption of the incident laser beam, the adsorption coefficients were measured by using an ultraviolet–visible (UV/vis) spectrophotometer at 532 nm to correct the incident and scattered laser strength. The fact that wheel-shaped anions were slowly oxidized helped to reduce the disturbing absorption (see below). The LLS measurement shows correspondingly that the scattered intensity continuously decreases owing to vesicle destruction. The smaller molybdates(vI), being decomposition products, do not contribute much to the scattered intensity. Only the final vesicles with the correct number of Mo(v) centres give the light-scattering signals observed. UV/vis spectrometry was used to determine the final nanowheel concentration based on the known extinction coefficient. The same supramolecular structures are of course also observed under exclusion of O₂, that is, in dark blue solutions.

Dynamic light scattering

DLS measures the intensity–intensity time correlation function by means of a BI-9000AT multi-channel digital correlator. The field correlation function $|g^{(1)}(\tau)|$ was analysed by the constrained regularized CONTIN method¹⁰, to yield information on the distribution of the characteristic linewidth Γ from $|g^{(1)}(\tau)| = \int G(\Gamma)e^{-\Gamma\tau}d\Gamma$. The normalized distribution function of the characteristic linewidth, $G(\Gamma)$, so obtained, can be used to determine an average apparent translational diffusion coefficient, $D_{app} = \Gamma/q^2$. The hydrodynamic radius R_h is related to D via the Stokes–Einstein equation: $R_h = kT/(6\pi\eta D)$ where k is the Boltzmann constant and η the viscosity of the solvent at temperature T . From DLS measurements, we can obtain the particle-size distribution in solution from a plot of $\Gamma G(\Gamma)$ versus R_h . The R_h of the particles is obtained by extrapolating $R_{h,app}$ to zero solute concentration and zero scattering angle. The R_h value of the $[Mo_{154}]/[Mo_{152}]$ vesicles does not show any obvious angular dependence, suggesting a spherical nature of the aggregates.

Transmission electron microscopy

Samples were prepared by pipetting 5 μ l of solution onto a freshly glow discharged carbon-coated electron microscopy grid, and after blotting excess liquid, 5 μ l of 2% uranyl acetate

aqueous solution was used to negatively stain the sample. The negative-stain method provides better sample contrast while keeping the structural details to around 1 nm resolution. Microscopy observation was performed in a Jeol JEM-1200EX microscope. Images were recorded on Kodak SO-163 films and developed with standard procedure. Negative films were digitized with a Nikon Super Coolscan 8000 scanner. The image-processing software SPIDER http://www.wadsworth.org/spider_doc/spider/docs/master.html was used to generate the power spectra.

Received 17 February; accepted 2 September 2003; doi:10.1038/nature02036.

- Evans, D. F. & Wennerström, H. *The Colloidal Domain: Where Physics, Chemistry, Biology, and Technology Meet* 2nd edn (Wiley, Chichester, 1999).
- Müller, A. *et al.* Rapid and simple isolation of the crystalline molybdenum-blue compounds with discrete and linked nanosized ring-shaped anions: $Na_{15}[Mo_{126}Mo_{28}O_{462}H_{14}(H_2O)_{70}]_{0.5}[Mo_{124}Mo_{28}O_{457}H_{14}(H_2O)_{68}]_{0.5} \cdot ca.400H_2O$ and $Na_{22}[Mo_{118}Mo_{28}O_{442}H_{14}(H_2O)_{58}] \cdot ca.250H_2O$. *Z. Anorg. Allg. Chem.* **625**, 1187–1192 (1999).
- Cronin, L., Diemann, E. & Müller, A. in *Inorganic Experiments* 2nd edn (ed. Woollins, J. D.) 340–346 (Wiley-VCH, Weinheim, 2003).
- Müller, A. & Serain, C. Soluble molybdenum blues—“des Pudels Kern”. *Acc. Chem. Res.* **33**, 2–10 (2000).
- Müller, A. *et al.* Hierarchic patterning: architectures far beyond “giant molecular wheels”. *Chem. Commun.* 1928–1929 (2001).
- Scheele, C. W. in *Sämtliche Physische und Chemische Werke* (ed. Hermstädt, D. S. F.) Vol. 1, 185–200 (Martin Sändig oHG, Niederwalluf/Wiesbaden, 1971) [reprint from 1793 original].
- Berzelius, J. J. Beitrag zur näheren Kenntniss des Molybdäns. *Poggend. Ann. Phys. Chem.* **6**, 369–392 (1826).
- Müller, A., Kögerler, P. & Kuhlmann, C. A variety of combinatorially linkable units as disposition: from a giant icosahedral Keplerate to multi-functional metal-oxide based network structures. *Chem. Commun.* 1347–1358 (1999).
- Rösler, H. J. *Lehrbuch der Mineralogie* 418 (Deutscher Verlag für Grundstoffindustrie, Leipzig, 1991).
- Provencher, S. W. CONTIN: A general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. *Comput. Phys. Commun.* **27**, 229–242 (1982).
- Hiemenz, P. C. & Rajagopalan, R. *Principles of Colloid and Surface Chemistry* Ch. 5 (Marcel Dekker, New York, 1997).
- Zhou, S. *et al.* Spherical bilayer vesicles of fullerene-based surfactants in water: a laser light scattering study. *Science* **291**, 1944–1947 (2001).
- Polarz, S., Smarsly, B. & Antonietti, M. Colloidal organization and clusters: self-assembly of polyoxometallate-surfactant complexes towards three-dimensional organized structures. *ChemPhysChem* **2**, 457–461 (2001).
- Müller, A. *et al.* Generation of cluster capsules (I_h) from decomposition products of a smaller cluster (Keggin - T_d) while surviving ones get encapsulated: species with core-shell topology formed by a fundamental symmetry-driven reaction. *Chem. Commun.* 657–658 (2001).
- Müller, A. The beauty of symmetry. *Science* **300**, 749–750 (2003).
- Caspar, D. & Klug, A. Physical principles in construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24 (1962).

17. Kurth, D. G., Volkmer, D., Ruttorf, M., Richter, B. & Müller, A. Ultrathin composite films incorporating the nanoporous isopolyoxomolybdate "Keplerate" $(\text{NH}_4)_{42}[\text{Mo}_{132}\text{O}_{372}(\text{CH}_3\text{COO})_{36}(\text{H}_2\text{O})_{72}]$. *Chem. Mater.* **12**, 2829–2831 (2000).
18. Polarz, S., Smarsly, B., Göltner, C. & Antonietti, M. The interplay of colloidal organization and oxo-cluster chemistry: polyoxometalate-silica hybrids—materials with a nanochemical function. *Adv. Mater.* **12**, 1503–1507 (2000).
19. Müller, A., Shah, S. Q. N., Bögge, H. & Schmidtman, M. Molecular growth from a Mo_{176} to a Mo_{248} cluster. *Nature* **397**, 48–50 (1999).
20. Müller, A., Toma, L., Bögge, H., Schmidtman, M. & Kögerler, P. Synergetic activation of "silent receptor" sites leading to a new type of inclusion complex: integration of a 64-membered ring comprising K^+ and SO_4^{2-} ions into a molybdenum oxide-based nanoobject. *Chem. Commun.* 2000–2001 (2003).
21. Liu, T. B. Supramolecular structures of polyoxomolybdate-based giant molecules in aqueous solution. *J. Am. Chem. Soc.* **124**, 10942–10943 (2002).
22. Liu, T. B. An unusually slow self-assembly of inorganic ions in dilute aqueous solution. *J. Am. Chem. Soc.* **125**, 312–313 (2003).

Acknowledgements We thank M. Schmidtman for the generation of Fig. 1d. T.L. acknowledges support of this work by the US Department of Energy, Division of Materials Science, and A.M. correspondingly, the Deutsche Forschungsgemeinschaft, the European Union, the Fonds der Chemischen Industrie, and the Volkswagen-Stiftung. H.L. acknowledges the support of the LDRD fund from BNL.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.L. (liu@bnl.gov) or A.M. (a.mueller@uni-bielefeld.de).

Interaction of sea water and lava during submarine eruptions at mid-ocean ridges

Michael R. Perfit¹, Johnson R. Cann², Daniel J. Fornari³, Jennifer Engels⁴, Deborah K. Smith³, W. Ian Ridley⁶ & Margo H. Edwards³

¹Department of Geological Sciences, University of Florida, Gainesville, Florida 32611, USA

²Department of Earth Sciences, University of Leeds, Leeds LS2 9JT, UK

³Woods Hole Oceanographic Institution, Geology and Geophysics Department, Woods Hole, Massachusetts 02543, USA

⁴Department of Geology and Geophysics, ⁵Hawaii Institute of Geophysics and Planetology, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, Honolulu, Hawaii 96822, USA

⁶US Geological Survey, Mineral Resources Team, Denver, Colorado 80225, USA

Lava erupts into cold sea water on the ocean floor at mid-ocean ridges (at depths of 2,500 m and greater), and the resulting flows make up the upper part of the global oceanic crust¹. Interactions between heated sea water and molten basaltic lava could exert significant control on the dynamics of lava flows and on their chemistry. But it has been thought that heating sea water at pressures of several hundred bars cannot produce significant amounts of vapour^{2–5} and that a thick crust of chilled glass on the exterior of lava flows minimizes the interaction of lava with sea water. Here we present evidence to the contrary, and show that bubbles of vaporized sea water often rise through the base of lava flows and collect beneath the chilled upper crust. These bubbles of steam at magmatic temperatures may interact both chemically and physically with flowing lava, which could influence our understanding of deep-sea volcanic processes and oceanic crustal construction more generally⁶. We infer that vapour formation plays an important role in creating the collapse features that characterize much of the upper oceanic crust and may accordingly contribute to the measured low seismic velocities in this layer.

Lobate and sheet lava flows erupted at axes of intermediate- and fast-spreading mid-ocean ridges (MORs) can extend for kilometres from their eruptive fissures as uppermost (~0–500 m) oceanic crust is constructed¹. MOR lava flows commonly contain depressions where thin crusts, 2–10 cm thick, forming flow surfaces, have collapsed into cavities beneath (see, for example, refs 6–11). Tabular pieces of lava crust often accumulate on the floor of the depressions; alternatively, the roof of the flow may remain in place, supported by lava pillars (Fig. 1a). The common occurrence of lava pillars in this setting is explained by the passage of large volumes of heated sea water up through an advancing and inflating lava flow or pond^{12–15}.

Sea-floor lava crusts have 1–2-cm-thick glassy margins on their exterior surfaces formed during the quenching of molten lava by cold (~2–4 °C) sea water, but frequently have thinner, partly glassy margins on their underside surface. Delicate drip structures and a variety of other surface morphologies are commonly observed on the underside of crusts, but not on the exterior surfaces. The features include: (1) mesocrystalline to glassy rinds a few milli-

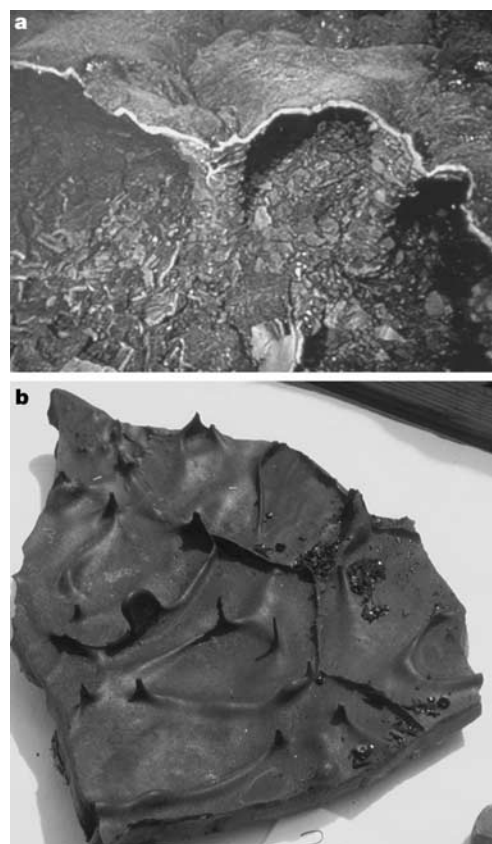


Figure 1 Macroscopic features of mid-ocean-ridge basalts indicative of magma–sea water interaction. **a**, Side-view into a deep-sea lobate lava flow on the East Pacific Rise crest near 9° 50' N at 2,510 m depth. Two lava pillars in the central portion of the photograph are supporting the bowed-up upper crust of the flow that is ~4–10 cm thick. Distance between flow surface and the floor is ~1 m. Platy talus from collapse of portions of the lava crust cover the sea floor in the foreground. Width across photograph taken from the submersible *Alvin* is ~4 m. (Photograph courtesy of Woods Hole Oceanographic Institution, Deep Submergence Operations Group.) **b**, Interior of a deep-sea lobate lava crust (concave side that faces down) from the East Pacific Rise crest at 2,510 m depth showing delicate drip structures and well-developed septa that separate areas where bubbles and cavities of briny vapour at magmatic temperature existed during formation of the drips. Width across photograph is 20 cm, individual drips are ~2–4 cm long. Although the surface appears to be smooth and glassy, optical microscopy and SEM images (see Figs 2 and 3) show a variety of textures and minerals comprise these surfaces.

metres thick with local concentrations of microvesicles lined with crystals; (2) drip-like protrusions up to 4 cm in length; and (3) septa <1 mm to 5 mm thick with flange to cusped morphology that often link drips or appear to be remnants of walls of vapour bubbles (Fig. 1b). All of these features can be explained if a low-density fluid/vapour cavity existed beneath the brittle upper crust when the lava flow was active. Temperatures within the cavities must have been substantially above the basalt solidus temperature to allow drips, septa and flanges to form on the cavity walls. Drip-like features have similar morphologies to lava stalactites and lava straws that form at magmatic temperatures on the roofs of subaerial lava tubes¹⁶. The drips have previously been observed, but their implications have

only recently been recognized^{1,6,14}. The cusped septa have features similar to bubble-walls or -shells formed during hydromagmatic eruptions on seamounts and submarine portions of oceanic islands^{2,5,17}.

Analysis of the cavity walls, using a scanning electron microscope (SEM) with an energy-dispersive X-ray spectrometer (EDX), confirms the near-magmatic temperatures within fluid/vapour cavities. Unusual silicate mineral textures on the underside of crusts and in associated pipe vesicles include open networks of plagioclase and almost pure forsterite associated with calcium-rich and calcium-poor pyroxenes, titanomagnetite and sulphide that line cavity walls and vugs (Fig. 2 and Supplementary Information). Although all of these minerals are expected to crystallize from a basaltic melt, the open, latticework textures and presence of pure forsterite in one instance and albitic plagioclase in another cavity are not consistent with simple cooling of a basaltic melt. Our identification of a thin (<5 mm) holocrystalline rind of elongate plagioclase and clinopyroxene with minor titanomagnetite and rare amphibole (Fig. 2 and Supplementary Information) on the underside of one thin crust from the East Pacific Rise suggests crystallization at the interface between a basaltic melt and a vapour phase. The presence of a vapour phase is supported by other unusual textures, such as encrustations of micro-fibrous tubes and whiskers, reticulate thin

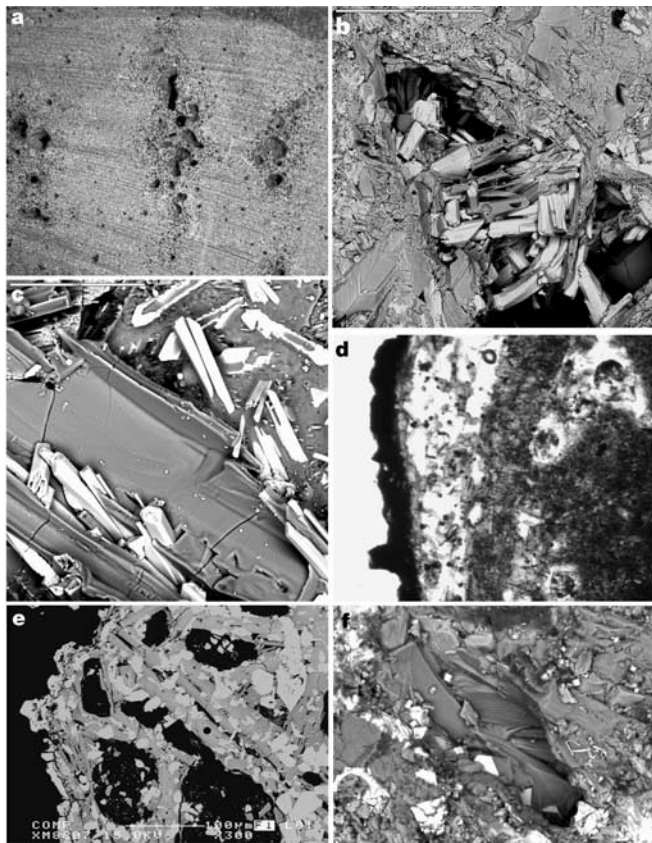


Figure 2 Various images of textures and minerals in deep-sea basalt associated with lava-vapour interaction. All samples are from the East Pacific Rise at a depth of approximately 2,500 m. **a**, Cut surface of lava flow showing pipe vesicles near bottom of flow. Top of flow is towards top of photograph. Width of photograph, ~2 cm. **b**, SEM image of the interior of one of the vesicles in **a** primarily showing plagioclase ($An_{54}-An_6$) and clinopyroxene radiating into the void space. Width of photograph, 400 μ m. **c**, Magnified SEM view of **b** showing plagioclase (dark grey), with smaller crystals of clinopyroxene (light grey), and titanomagnetite (white, top right) with rare olivine. Width of photograph, 100 μ m. **d**, Photomicrograph of bottom crust of basalt ALV 2697-5 exhibiting coarse intergrowths of plagioclase, clinopyroxene, titanomagnetite and olivine in a zone that is highly vesicular (white irregular zones near edge). The dark outer rim on the left is a thin Mn-coating. Towards the interior of the thin section (to the right of the photograph) the lava becomes less crystalline and is less vesicular. In some places along the edge, crystals are perpendicular to the crust surface exhibiting a comb-like texture. Thickness of lava shown is ~10 mm. **e**, Back-scattered electron (BSE) image of same area shown in **d** illustrating textural relationships between elongate plagioclase (dark grey) and clinopyroxene (light grey) along the microvesicular underside of crust. Additional BSE images and analyses of phases are presented in Supplementary Table 2. Width of image, 400 μ m. **f**, SEM image of microvesicular zone on outside edge of (left side) sample shown in **d** and **e** showing plagioclase (dark grey), clinopyroxene (lighter grey) and titanomagnetite (white). Width of image, 100 μ m.

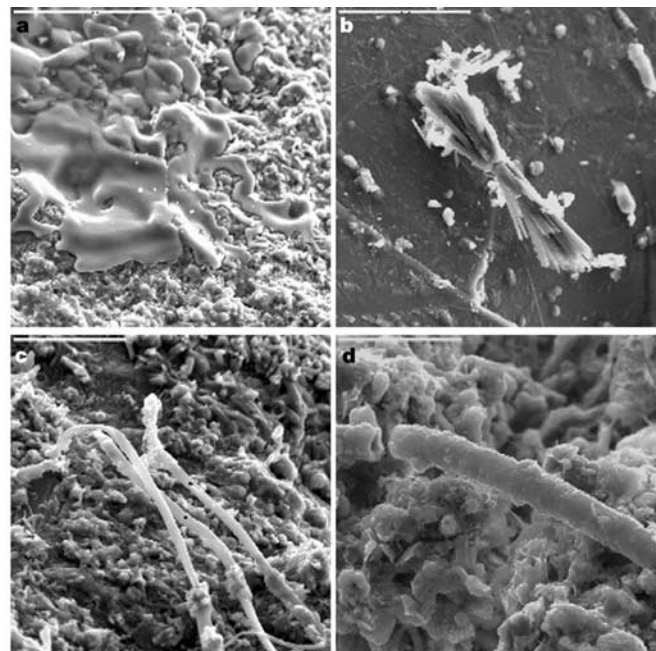


Figure 3 Microscopic features observed by SEM analysis of undersides of basalt crusts collected at the East Pacific Rise near 9° 50' N from a depth of 2,510 m. **a**, Pure NaCl covering secondary minerals on the surface of a lava drip such as those seen in Fig. 1. The very smooth structure of the salt (even at $\times 70,000$) is consistent with the hypothesis that it was molten and flowed over the other minerals. Width of photograph, 100 μ m. **b**, Classic bow-tie structure in a chemically complex secondary silicate containing an unusually high concentration of Cu (~7.5 wt%). Width of photograph, 50 μ m. **c**, Three filaments lying on secondary minerals on a drip surface. Note that each filament has further secondary mineral deposits on its surface. These filaments have 8% Cl, 7% S, 10% MnO, 16% FeO and very low SiO_2 (14%) and other constituents. All of the EDS analyses of secondary minerals give low totals (80–90 wt%) suggesting that they are all hydrous minerals. Width of photograph, 70 μ m. **d**, A higher-resolution image of a filament straw on a lava drip surface showing the central channel that often typifies these structures. This particular filament contains >60% MnO + FeO. Width of photograph, 20 μ m. Many of the observed minerals extend outward from partially crystalline lava surfaces that are either pitted, microvesicular, or have an unusual lattice-work texture (see Fig. 2).

plates and blades, and a variety of euhedral secondary minerals (Fig. 3). Locally, these minerals are coated with a smooth layer of sodium chloride, the texture being suggestive of molten salt that would require deposition above its melting point of 801 °C (Supplementary Table 1). Chemical analysis by EDX of a variety of these minerals (including Ca sulphate and Fe sulphate) on cavity surfaces shows many of them to have complex compositions. They can contain several per cent of chlorine and sulphur and high concentrations of copper and manganese.

Such crystal forms and minerals have not previously been documented on unaltered, glassy sea-floor lavas, and require unusual processes to form them. First, an enriched source of some elements, for example, Cl and S, is required. An obvious choice is sea water. Second, the high concentrations of some other elements (for example, Cu, Mn and Fe) are consistent with selective removal from a basaltic melt into a vapour/fluid phase at near-magmatic temperatures. The mineral assemblages and their textures indicate formation from a silicate melt in equilibrium with a vapour phase at magmatic temperatures¹⁸. We have not observed these features on the glassy outer surfaces of lava crusts exposed to cold sea water.

The presence of a vapour phase at depths >2,500 m has been proposed to explain mineral deposition in small cavities, drips, bubble shards and some volcanoclastic deposits^{1–5,14,17–20}, but this vapour phase has generally been thought to come from exsolution of magmatic volatiles^{4,5,18–20}. However, volatile contents in MOR basalts are too low and hydrostatic pressures are too great (~25 MPa at 2,500 m depth) for exsolution of H₂O and CO₂ to produce more than ~5 vol.% of vesicles from volatiles dissolved in the melt^{1,17,21,22}. Furthermore, the diffusive growth of magmatic vesicles in cooling submarine lava flows is too slow to form and accumulate into large bubbles or pockets¹⁷. Although it is possible that large amounts of CO₂ may be emitted with the lava at the eruptive fissure^{4,5}, as happens during basaltic eruptions on land, cavities and collapse features are seen in lavas that have flowed well away (up to a few kilometres) from the MOR axis—sites where excess CO₂ is unlikely to be present^{1,6,9}.

We argue that the vapour phase is derived from sea water interacting with MOR lava flows as they advance over irregular sea floor. The principal obstacle to such a view in the past has been the interpretation that heating sea water to high temperatures (>400 °C) at pressures as great as those at the deep-sea floor does not form substantial volumes of steam^{2–5}. The NaCl–H₂O phase diagram at 30 MPa (refs 23, 24) shows that sea water boils at ~400 °C, and at 480 °C the system is 95% vapour and 5% fluid (Supplementary Table 1). Between 480 °C and ~900 °C the system is a mixture of vapour and concentrated brine or molten NaCl. Above ~900 °C a single-phase vapour of seawater salinity is present occupying a volume 20 times greater than the cold fluid.

We can also calculate the rate at which steam would be produced below a lava flow, using standard equations of heat conduction through a hot interface. For example, if one-third of the heat conducted through the base of the flow produces steam that can rise through the flow, after 5 min a layer of water 0.5 cm thick will have vaporized to a layer of steam 10 cm thick. After 50 min a layer of water 2 cm thick will have vaporized to produce a 40-cm-thick layer of steam. Lava pillars provide striking evidence for ascent of large volumes of sea water and vapour through MOR flows^{13–14}. We believe that smaller volumes of vapour can be incorporated within sheet and lobate flows, as they advance over young sea floor that is extensively fractured and contains water-filled voids. Bubbles of vapour will ascend through low-viscosity magma towards the upper surface, where they initially can collect to form centimetre-sized cavities directly below the chilled glassy outer crust of the flow. Coalescence of small bubbles results in the formation of a vapour cavity that separates the inside of the upper crust from molten magma below, thus preventing further inward growth of the flow's

upper crust. Glassy crust a few centimetres thick takes from a few minutes to a few tens of minutes to develop on the top surface of an advancing submarine lava flow^{25–27}. Thus vapour bubbles must have started to grow soon after the flow front passed. As the lava flow cools and solidifies, the steam-filled cavity starts to cool, and pressure within the cavity decreases. This imposes an increasingly large pressure differential across the crust above the bubble, until eventually it fractures and founders. Rapid inflow of sea water quenches the steam in the bubble (and drips that have formed), and fills the cavity with sea water. The result is the formation of collapse pits that are common in submarine lava flows.

Evidence for transport of vapour through a flow as bubbles or tubes comes from pipe vesicles and spiracles preserved in ancient flows and modern MOR lavas (refs 6, 28; see also Fig. 2a). Sea water may also be incorporated into a flow from its top surface. Pressure within a mature lava flow or lava tube system on the sea floor will decrease if the lava flow rate slackens, or if lava withdraws from the tube^{6,15}. The presence of drips and septa provides evidence that magma flowed in a high-temperature vapour-filled cavity, and that the overall magma–vapour system must have been close to the basalt solidus temperature (>1,070 °C)²⁹.

What effect will the presence of a large volume of hot water vapour have on the dynamics of a lava flow? The presence of vapour bubbles will reduce the friction between the body of the flow and its roof, floor and walls, allowing flows to extend farther across the sea floor even in areas where sea-floor gradients are low. The potential effect of vapour in sea-floor lava tubes is also important. In subaerial lava tubes, waxing and waning of magma flow can be facilitated by flow of air through 'skylights'²⁹. On the sea floor, there is compelling evidence for long-term use of single lava tubes for transport of large volumes of magma¹⁰. In a deep-sea tube, the presence of a buffer of steam above the lava would allow a flow to be maintained despite variation in lava effusion rate.

Lava chemistry may also be affected by the hot briny vapour. MOR basalt is under-saturated with water at 25–30 MPa (ref. 30). Rapid chilling of a lava flow exterior prevents water being absorbed by, or diffusing into, the magma from outside, but the same would not apply to high-temperature steam inside a flow. Though diffusion would not be fast across the magma/gas interface, the skin of magma in contact with the steam bubble will absorb some water. Sodium chloride and other seawater solutes could also enter the lava in this way. Drops of highly saline brine in equilibrium with steam would sink to the bottom of the steam bubble, and come into direct contact with magma. The scale of the influence of assimilated sea water on overall chemistry of ocean-floor basalts would depend on the extent of mixing of contaminated lava with the bulk of the flow. Many unaltered MOR basalt glasses have high Cl/K ratios, suggesting some type of magma–seawater/brine interaction²² or assimilation of seawater-altered crust³¹. Identifying these different processes and their magnitudes are critical to our understanding of MOR petrogenetic processes.

We conclude that the formation of steam-filled cavities in lava erupting at MORs explains the formation of ubiquitous collapse features in shallow oceanic crust, and that seawater–lava interaction can affect the primary composition of MOR basalt. Vapour formation during sea-floor eruptions is also likely to facilitate the distribution of lava along and across the ridge crest, forming a carapace hundreds of metres thick covering more than half the globe. □

Received 2 July; accepted 2 September 2003; doi:10.1038/nature02032.

1. Perfit, M. R. & Chadwick, W. W. in *Faulting and Magmatism at Mid-Ocean Ridges* (eds Buck, R. W., Delaney, P. T., Karson, J. A. & Lagabriele, Y.) AGU Monograph 106 59–115 (American Geophysical Union, Washington DC, 1999).
2. Batiza, R. & White, J. D. L. in *Encyclopedia of Volcanology* (ed. Sigurdsson, H.) 383–402 (Academic, San Diego, 2000).
3. Haymon, R. M. *et al.* Volcanic eruption of the mid-ocean ridge along the East Pacific Rise crest at 9 degrees 45–52' N: Direct submersible observations of seafloor phenomena associated with an eruption

Sulphide mining by the superextensile foot of symbiotic thyasirid bivalves

Suzanne C. Dufour & Horst Felbeck

Marine Biology Research Division, Scripps Institution of Oceanography, 9500 Gilman Drive, La Jolla, California 92093-0202, USA

In a symbiotic association between an invertebrate host and chemoautotrophic bacteria, each partner has different metabolic requirements, and the host typically supplies the bacteria with necessary reduced chemicals (sulphide or methane). Some combination of anatomical, physiological and behavioural adaptations in the host often facilitates uptake and transport of reduced chemicals to the symbionts^{1–4}. We have studied five species of bivalve molluscs of the family Thyasiridae (that is, thyasirids) three of which harbour chemoautotrophic bacteria. Here we show that the symbiotic bivalves extend their feet to form elongated and ramifying burrows in the sediment, most probably to gain access to reduced sulphur. Closely related bivalves (including some thyasirid species) without bacterial symbionts show no comparable foot extension behaviour. The length and number of burrows formed by chemosymbiotic thyasirids are related to the concentration of hydrogen sulphide in the sediment. The burrows are formed by the foot of each bivalve, which can extend up to 30 times the length of the shell, and may be the most extreme case of animal structure elongation documented to date.

To maintain their symbiosis, bivalves and other invertebrates associated with sulphur-oxidizing bacteria must provide their symbionts with reduced sulphur, while still having access to oxygen. Accordingly, chemosymbiotic animals occur near oxic/anoxic disequilibrium zones, such as hydrothermal vents or reducing sediments. Whereas some chemosymbiotic invertebrates periodically bathe in dissolved sulphide, others, such as thyasirid and lucinid bivalves, are found in sediments where ambient sulphide levels are very low or undetectable^{5,6}. These bivalves presumably have adaptations for obtaining sulphide in these particular environments.

Observations of symbiotic thyasirids and lucinids burrowing at the perimeter of glass aquaria suggest a mechanism for sulphide acquisition in these species. Burrows detected underneath the animals⁶ were interpreted as being conduits for sulphide, comparable to the burrows of *Solemya velum*, through which sulphide-laden water is said to be pumped into the mantle cavity⁷. Lucinacean burrows could be dug by the foot, which forms the mucus-lined anterior inhalant tube⁸ and was seen to extend up to five times the shell length in lucinids⁹, and about fifteen times the shell length in *Thyasira flexuosa*⁷.

Alternatively, locomotion in the sediment may differ between related bivalves with and without symbionts, revealing adaptations

Table 1 Burrows formed by chemosymbiotic thyasirids after establishment

	Mean number of burrows	Mean cumulative burrow length (cm)	Maximum burrow length (cm)
High sulphide (14.3 ± 1.9 μmol l ⁻¹)*	1.3 ± 0.2*	3.6 ± 0.5*	5.1
Low sulphide (6.1 ± 1.0 μmol l ⁻¹)*	2.5 ± 0.3*	6.9 ± 0.9*	13.9

Values are averages for weeks 2 and 3; means are shown ± s.e.m. Asterisks denote statistical significance ($P \leq 0.001$) in *t*-tests comparing variables from the high and low sulphide treatments.

- event in April, 1991. *Earth Planet. Sci. Lett.* **119**, 85–101 (1993).
- Head, J. W. & Wilson, L. Deep submarine pyroclastic eruptions: Theory and predicted landforms and deposits. *J. Volcanol. Geotherm. Res.* **121**, 155–193 (2003).
 - Clague, D. A., Davis, A. S., Bischoff, J. L., Dixon, J. E. & Geyer, R. Lava bubble-wall fragments formed by submarine hydrovolcanic explosions on Loihi Seamount and Kilauea Volcano. *Bull. Volcanol.* **61**, 437–449 (2000).
 - Engels, J., Edwards, M. H., Fornari, D. J., Perfit, M. R. & Cann, J. R. A new model for submarine volcanic collapse formation. *Geochem. Geophys. Geosyst.* **4**, 1078 (DOI:10.2939/2003GC000560) 18 September 2003.
 - Ballard, R. D., Holcomb, R. T. & van Andel, T. H. The Galapagos Rift at 86°W: Sheet flows, collapse pits, and lava lakes of the rift valley. *J. Geophys. Res.* **84**, 5407–5422 (1979).
 - Embley, R. W. & Chadwick, W. W. Volcanic and hydrothermal processes associated with a recent phase of seafloor spreading at the northern Cleft segment; Juan de Fuca Ridge. *J. Geophys. Res.* **99**, 4741–4760 (1994).
 - Fornari, D. J., Haymon, R. M., Perfit, M. R., Gregg, T. K. P. & Edwards, M. H. Geological characteristics and evolution of the axial zone on fast spreading mid-ocean ridges: Formation of an axial summit trough along the East Pacific Rise, 9°–10° N. *J. Geophys. Res.* **103**, 9827–9855 (1998).
 - Smith, D. K. & Cann, J. R. Constructing the upper crust of the Mid-Atlantic Ridge: A reinterpretation based on the Puna Ridge, Kilauea Volcano. *J. Geophys. Res.* **104**, 25379–25399 (1999).
 - Sinton, J. et al. Volcanic eruptions on MORs: new evidence from the superfast-spreading EPR, 17°–19°S. *J. Geophys. Res.* **107** (DOI:10.1029/2000JB000090) (2000).
 - Francheteau, J., Juteau, T. & Rangan, C. Basaltic pillars in collapsed lava-pools on the deep ocean floor. *Nature* **281**, 209–211 (1979).
 - Gregg, T. K. P. & Chadwick, W. W. Submarine lava-flow inflation: a model for the formation of lava pillars. *Geology* **24**, 981–984 (1996).
 - Gregg, T. K. P., Fornari, D. J., Perfit, M. R., Ridley, W. I. & Kurz, M. D. Using submarine lava pillars to record MOR eruption dynamics. *Earth Planet. Sci. Lett.* **178**, 195–214 (2000).
 - Chadwick, W. W. Jr Quantitative constraints on the growth of submarine lava pillars from a monitoring instrument that was caught in a lava flow. *J. Geophys. Res.* (in the press).
 - Peterson, D. W. & Swanson, D. A. Observed formation of lava tubes. *Stud. Speleol.* **2**, 209–222 (1974).
 - Maicher, D. & White, J. D. L. The formation of deep-sea Limu o Pele. *Bull. Volcanol.* **63**, 482–496 (2001).
 - Sharapov, V. N., Pavlov, A. L., Akimsev, V. A. & Zhmodik, A. S. Physicochemical conditions of mineral deposition from magmatic gases in basalts of the mid-ocean ridges. *Geol. Ore Deposits* **43**, 76–87 (2001).
 - Fouquet, Y. et al. Extensive volcanoclastic deposits at the mid-Atlantic Ridge axis: Results of deep-water basaltic explosive activity? *Terra Nova* **10**, 280–286 (1998).
 - Tribble, G. W. Underwater observations of active lava flows from Kilauea volcano, Hawaii. *Geology* **19**, 633–636 (1991).
 - Dixon, J. E., Stolper, E. & Delaney, J. R. Infrared spectroscopic measurements of CO₂ and H₂O in Juan de Fuca basaltic glasses. *Earth Planet. Sci. Lett.* **90**, 87–104 (1988).
 - Le Roux, P. J., Shirey, S. B., Hauri, E. H., Perfit, M. R. & Mock, T. Degassing and preliminary assimilation histories of selected on- and off-axis EPR MORB glasses. *Goldschmidt Conf. Abstr. Geochem. Soc. A* **437** (2002).
 - Sourirajan, S. & Kennedy, G. C. The system H₂O–NaCl at elevated temperatures and pressures. *Am. J. Sci.* **260**, 115–141 (1962).
 - Berndt, M. E. & Seyfried, W. E. Calibration of Br/Cl fractionation during sub-critical phase separation of seawater; Possible halite at 9 to 10 degrees N, East Pacific Rise. *Geochim. Cosmochim. Acta* **61**, 2849–2854 (1997).
 - Griffiths, R. W. & Fink, J. H. Solidification and morphology of submarine lavas: A dependence on extrusion rate. *J. Geophys. Res.* **97**, 19729–19737 (1992).
 - Gregg, T. K. P. & Fink, J. H. Quantification of submarine lava-flow morphology through analog experiments. *Geology* **23**, 73–76 (1995).
 - Klingelhofner, E., Hort, M., Kumpel, H.-J. & Schmincke, H.-U. Constraints on the formation of submarine lava flows from numerical model calculations. *J. Volcanol. Geotherm. Res.* **92**, 215–229 (1999).
 - Watters, A. C. Determining direction of flow in basalts. *Am. J. Sci.* **A 258**, 350–366 (1960).
 - Hon, K., Kauahikaua, J., Denlinger, R. & Mackay, K. Emplacement and inflation of pahoehoe sheet flows: Observations and measurements of active lava flows on Kilauea Volcano. *Hawaii. Geol. Soc. Am. Bull.* **106**, 351–370 (1994).
 - Dixon, J. E., Stolper, E. M. & Holloway, J. R. An experimental study of water and carbon dioxide solubilities in mid ocean ridge basaltic liquids. 1. Calibration and solubility models. *J. Petrol.* **36**, 1607–1631 (1995).
 - Michael, P. J. & Cornell, W. C. Influence of spreading rate and magma supply on crystallization and assimilation beneath mid-ocean ridges: Evidence from chlorine and major-element chemistry of mid-ocean ridge basalts. *J. Geophys. Res.* **103**, 18325–18356 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the Deep Submergence Operations Group, and *Alvin* and *R/V Atlantis* crews at Woods Hole Oceanographic Institution for assistance in collecting these data. We also thank I. Jonasson for pointing out similar features on the Juan de Fuca Ridge. M. Smith, W. Chadwick, D. Clague, T. Gregg, M. Tivey and S. Humphris provided discussions and comments on the manuscript. J. E. Dixon and J. C. Fink provided reviews. J.R.C. was supported in part by internal funds of Woods Hole Oceanographic Institution. W.I.R. publishes with permission of the Director, US Geological Survey. This work was supported by the National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.R.P. (perfit@geology.ufl.edu.).

for seeking out or avoiding sulphide. The family Thyasiridae, which has species with and without chemoautotrophic symbionts, can be separated into two groups: those with upright (dorso-ventrally extended) shells and those with elongate (antero-posteriorly extended) shells¹⁰. Thyasirids with upright shells seem better adapted for deep burrowing, not only because their shell shape offers less resistance to sediment penetration, but also because each usually has a long foot, capable of constructing an anterior inhalant tube¹⁰. In contrast, thyasirids with elongate shells (most of which have a shorter foot, with a heel) may live just beneath the sediment surface, and may be more motile than the first group¹⁰. Although these hypothesized differences in life habits were not studied in the context of chemosymbiosis, we have observed that known symbiotic thyasirids all have upright shells and an elongate foot.

To evaluate whether bivalves use different locomotory abilities, and/or form burrows as sulphide acquisition strategies, the behaviours of symbiotic and non-symbiotic thyasirids were compared. The living animals were observed burrowing in thin aquaria, which were X-rayed weekly. The experimental design also permitted a test of whether the burrowing behaviour was affected by the concentration of sulphide dissolved in the pore water: 'high' and 'low' sulphide treatments were used (see Methods). Five thyasirid species were studied: *Thyasira* (*Parathyasira*) *equalis*, *Thyasira obsoleta*, *Thyasira* (*Mendicula*) *ferruginea*, *Thyasira flexuosa* and *Thyasira sarsi*. On the basis of previous reports, *T. equalis*, *T. flexuosa* and *T. sarsi* have symbionts¹¹, whereas *T. obsoleta* and *T. ferruginea* do not (ref. 11, and S.C.D., manuscript in preparation).

Measurements of sulphide content in each aquarium confirmed that the sulphide levels in each treatment differed significantly (Table 1, $P \leq 0.001$). High sulphide content in samples (above $10 \mu\text{M}$) could, in most cases, be confirmed by smell. The average

sulphide content per aquarium increased with time during the experiment.

After thyasirids were placed at the sediment surface, most burrowed within a few days. After one week, all specimens were located within 3.5 cm under the sediment surface, roughly corresponding to the *in situ* depth distribution reported for similar-sized *T. flexuosa* and *T. sarsi*⁶. For many specimens, the resulting trace of this burrowing activity appears as broad bands on the X-radiographs (Fig. 1). Inhalant tubes connecting the anterior end of the animal with the sediment surface were visible next to some of the symbiotic thyasirids; others had perhaps not yet constructed such tubes.

Between weeks, most thyasirids changed position in the sediment. Many had penetrated 5–7 mm deeper from one week to the next, and a few had returned to the sediment surface. They were never found deeper than 3.5 cm, however. Thyasirids without symbionts remained within 1.5 cm from the sediment surface, and did not appear more motile than symbiotic thyasirids.

Burrows underneath thyasirid shells were visible on all micrographs, and were associated only with symbiotic species. The burrows consisted of one or more tracks, sometimes ramified, and not necessarily originating from the same point at the ventral edge of the bivalve (Fig. 1). These burrows most probably represent insertions of the foot within the sediment, with ramifications indicating that the foot was withdrawn to a certain point and re-extended in a different direction. After one week, some of the symbiotic thyasirids had formed one or a few burrows, which were up to 7.3 cm in length. The burrows were much longer and more numerous (at most, 11 per animal) after two and three weeks, reaching a maximum length of 13.9 cm (30 times the shell length). Individual burrows did not remain X-ray detectable for more than a week.

Striking differences were observed in the behaviour of thyasirids in low and high sulphide conditions, after two and three weeks. In the low sulphide treatment, the burrows were longer, and more numerous than in the high sulphide treatment (Fig. 2, Table 1). This implies that the foot extensions are a sulphide acquisition mechanism; where there is less sulphide in the sediment, the bivalves presumably need to mine a larger space to acquire this resource. It appears that the foot of chemosymbiotic thyasirids is able to sense sulphide and display chemotaxis: in a few aquaria, burrows reveal the foot making a sharp turn to access what seems to be a pocket of sulphide (a darker spot within the sediment, often surrounding a trapped gas bubble). The differing lengths and number of branches

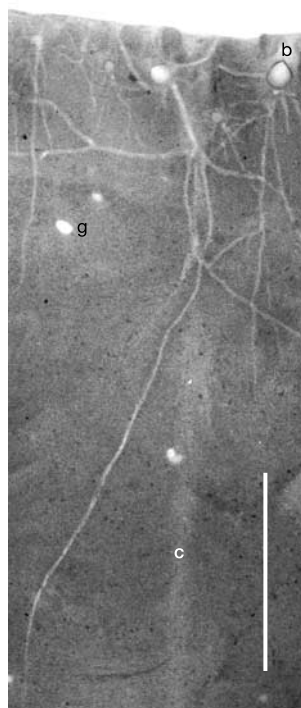


Figure 1 Partial X-radiograph of chemosymbiotic thyasirids burrowing in sediment two weeks after introduction. Traces above the animals show the paths taken during burial (labelled 'b'), and some anterior inhalant tubes, while traces below the animals represent sulphide mining burrows. All thyasirids shown here are chemosymbiotic. 'c' indicates trace of a sediment core taken the previous week; 'g' shows gas bubble trapped in the sediment. Scale bar, 5 cm.

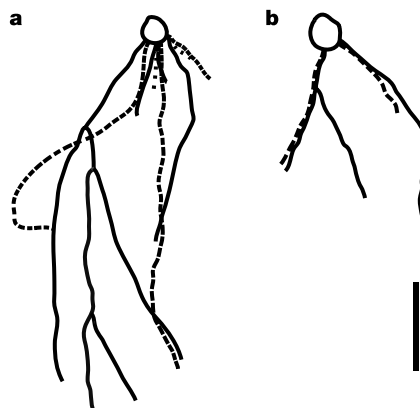


Figure 2 Typical structure of sulphide mining burrows formed by chemosymbiotic thyasirids in low and high sulphide conditions. **a**, Low sulphide; **b**, high sulphide. Long-dashed lines, burrows visible after one week; short-dashed lines, burrows visible after two weeks; lines, burrows visible after three weeks. Scale bar, 1 cm.

of burrows in high and low sulphide treatments provide further evidence for chemosensory abilities.

In thyasirids and lucinids, the foot consists of an outer layer of longitudinal and circular muscles, surrounding and traversing a spongy haemocoel⁸. The mechanism of foot extension in thyasirids is likely to have a hydrostatic component, with haemolymph being pumped by the heart from the body mass to the foot (the Keber valve stops haemolymph from returning to the heart¹²), resulting in an increase in foot length provided that the diameter of the foot remains constant or is reduced. Hydrostatic extension, although common in burrowing marine invertebrates, has not been reported to cause such a remarkable increase in length as seen in chemosymbiotic thyasirids (30 times the contracted length). Fully extended, the foot of a chemosymbiotic thyasirid must have an extremely thin, stretched-out epithelium and muscle layer. In this state, sulphide from pore water surrounding the burrow could easily diffuse to the haemal space of the foot. Then, sulphide could be transported to the gills and symbionts when the foot is retracted within the mantle cavity. Transport of sulphide from the foot to the gills has previously been proposed for the chemosymbiotic bivalves *Calyptogena magnifica* and *Calyptogena elongata*^{13,14}.

The sulphide mining behaviour could have been derived from the use of the foot in anterior inhalant tube formation, characteristic of Lucinaceans (lucinids, thyasirids and unguinids)⁸. On X-radiographs of the non-chemosymbiotic unguinid *Diplodonta notata*, some burrows are directed towards the sediment surface, whereas others lead deeper into the sediment, apparently having been misdirected¹⁵. Such accidental burrowing could have evolved into sulphide mining in chemosymbiotic thyasirids and lucinids, as it brought them in contact with pockets of sulphide.

Sulphide mining in thyasirids is not only a recent phenomenon: a fossilized thyasirid from the Middle Miocene, with a tunnel underneath measuring 15 times the shell length, has been discovered in Lower Austria¹⁶. The burrowing behaviour of thyasirids seems to have appeared early in the evolution of the family, whose fossil record goes back to the Cretaceous period¹⁷. Given the lack of a rigorous thyasirid phylogeny, the relationship between symbiotic and non-symbiotic thyasirids, and their relative order of appearance, are unclear. However, if sulphide mining existed in early thyasirids, then non-symbiotic thyasirids may have subsequently lost the elongated foot and mining behaviour by pedomorphosis, which was previously invoked to explain the simple gill structure and small size of many non-symbiotic thyasirids¹⁸. Chemosymbiotic thyasirids, on the other hand, have taken sulphide mining to the extreme, by pushing the limits of hydrostatic extension for the benefit of bacterial symbionts. □

Methods

Thyasirids were collected from the Raunefjord (*T. Parathyasira* *equalis*, *T. obsoleta*, *T. (Mendicula)* *ferruginea*) and from Dolviken (*T. flexuosa*, *T. sarsi*), near Bergen, Norway. Sediment collected by Van Veen grab from the Raunefjord site was passed through a 1-mm sieve and distributed into 12 plexiglass aquaria (dimensions: 25 cm × 25 cm × 12 mm; inner width, 6 mm). The aquaria were subdivided into two groups, in which the sediment settled for four days ('low sulphide') or four weeks ('high sulphide') before the introduction of bivalves at the sediment surface. The aquaria were kept in 60-litre tanks containing aerated sea water, replaced daily, and phytoplankton from the Raunefjord (average concentration, 6.4×10^6 cells ml⁻¹). The tanks were maintained in the dark, at 8 °C. After 7, 14 and 21 days, the aquaria were X-rayed (instrument settings: 65 kV, 50 mA). On the X-radiographs, the thyasirids could not always be identified to the species level. However, symbiotic and non-symbiotic thyasirids could be distinguished on the basis of size (the non-symbiotic species were consistently smaller). The day after X-rays were taken, a thin core of sediment was extracted from each aquarium and pore water from each 5-cm interval was isolated by centrifugation. The sulphide content in each fraction was determined using a colorimetric method¹⁹.

Received 24 July; accepted 25 September 2003; doi:10.1038/nature02095.

- Childress, J. J., Felbeck, H. & Somero, G. N. Symbiosis in the deep sea. *Sci. Am.* **256**, 114–120 (1986).
- Le Pennec, M., Beninger, P. G. & Herry, A. Feeding and digestive adaptations of bivalve molluscs to sulphide-rich habitats. *Comp. Biochem. Physiol.* **A 111**, 183–189 (1995).
- Julian, D., Gaill, F., Wood, E., Arp, A. J. & Fisher, C. R. Roots as a site of hydrogen sulfide uptake in the hydrocarbon seep vestimentiferan *Lamellibrachia* sp. *J. Exp. Biol.* **202**, 2245–2257 (1999).

- Giere, O., Conway, N. M., Gastrock, G. & Schmidt, C. 'Regulation' of gutless annelid ecology by endosymbiotic bacteria. *Mar. Ecol. Prog. Ser.* **68**, 287–299 (1991).
- Dando, P. R., Southward, A. J. & Southward, E. C. Chemoautotrophic symbionts in the gills of the bivalve mollusc *Lucinoma borealis* and the sediment chemistry of its habitat. *Proc. R. Soc. Lond. B* **227**, 227–247 (1986).
- Dando, P. R. & Southward, A. J. Chemoautotrophy in bivalve molluscs of the genus *Thyasira*. *J. Mar. Biol. Assoc. UK* **66**, 915–929 (1986).
- Seilacher, A. Aberrations in bivalve evolution related to photo- and chemosymbiosis. *Hist. Biol.* **3**, 289–311 (1990).
- Allen, J. A. On the basic form and adaptations to habitat in the Lucinacea (Eumellibranchia). *Phil. Trans. R. Soc. Lond. B* **241**, 421–484 (1958).
- Allen, J. A. Function of the foot in the Lucinacea (Eumellibranchia). *Nature* **171**, 1117–1118 (1953).
- Payne, C. M. & Allen, J. A. The morphology of deep-sea Thyasiridae (Mollusca: Bivalvia) from the Atlantic Ocean. *Phil. Trans. R. Soc. Lond. B* **334**, 481–562 (1991).
- Southward, E. C. Gill symbionts in thyasirids and other bivalve molluscs. *J. Mar. Biol. Assoc. UK* **66**, 889–914 (1986).
- Trueman, E. R., Brown, A. C. & Stenton-Dozey, J. Blood flow in a burrowing bivalve at pedal extension and retraction. *J. Moll. Stud.* **52**, 265–266 (1986).
- Arp, A. J., Childress, J. J. & Fisher, C. R. Metabolic and blood gas transport characteristics of the hydrothermal vent bivalve *Calyptogena magnifica*. *Physiol. Zool.* **57**, 648–662 (1984).
- Childress, J. J., Fisher, C. R., Favuzzi, J. A., Arp, A. J. & Oros, D. R. The role of a zinc-based, serum-borne sulphide-binding component in the uptake and transport of dissolved sulphide by the chemoautotrophic symbiont-containing clam *Calyptogena elongata*. *J. Exp. Biol.* **179**, 131–158 (1993).
- Stanley, S. M. Relation of shell form to life habits in the bivalvia (Mollusca). *Geol. Soc. Am. Mem.* **125**, 1–293 (1970).
- Zuschin, M., Mandic, O., Harzhauser, M. & Pervesler, P. Fossil evidence for chemoautotrophic bacterial symbiosis in the thyasirid bivalve *Thyasira michelottii* from the Middle Miocene (Badenium) of Austria. *Hist. Biol.* **15**, 123–134 (2001).
- McAlester, A. L. Evolutionary and systematic implications of a transitional ordovician lucinoid bivalve. *Malacologia* **3**, 433–439 (1966).
- Reid, R. G. B. & Brand, D. G. Sulfide-oxidizing symbiosis in lucinaceans: implications for bivalve evolution. *Veliger* **29**, 3–24 (1986).
- Gilboa-Garber, N. Direct spectrophotometric determination of inorganic sulfide in biological materials and in other complex mixtures. *Anal. Biochem.* **43**, 129–133 (1971).

Acknowledgements We thank J. B. C. Jackson, N. D. Holland and E. Kisfaludy for discussion and advice, and A. Aadnesen, P. Johannessen, T. Sorlie and T. Høisæter for help in collecting thyasirids and running experiments at the University of Bergen Biological Station, Esprengend. We are grateful to O. K. Dalland and A. Lerheim for allowing the use of X-ray equipment at the veterinary clinic at Stend, Norway. This work was supported by the Scripps Institution of Oceanography graduate department, the Baxter and Alma Ricard Foundation, and the National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.C.D. (sdoufour@ucsd.edu).

Governance and the loss of biodiversity

R. J. Smith¹, R. D. J. Muir¹, M. J. Walpole¹, A. Balmford² & N. Leader-Williams¹

¹Durrell Institute of Conservation and Ecology, University of Kent, Canterbury, Kent CT2 7NS, UK

²Conservation Biology Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

Most of the world's biodiversity occurs within developing countries that require donor support to build their conservation capacity¹. Unfortunately, some of these countries experience high levels of political corruption², which may limit the success of conservation projects by reducing effective funding levels and distorting priorities³. We investigated whether changes in three well surveyed and widespread components of biodiversity were associated with national governance scores and other socioeconomic measures. Here we show that governance scores were correlated with changes in total forest cover, but not with changes in natural forest cover. We found strong associations between governance scores and changes in the numbers of African

elephants and black rhinoceroses, and these socio-economic factors explained observed patterns better than any others. Finally, we show that countries rich in species and identified as containing priority areas for conservation have lower governance scores than other nations. These results stress the need for conservationists to develop and implement policies that reduce the effects of political corruption and, in this regard, we question the universal applicability of an influential approach to conservation that seeks to ban international trade in endangered species.

Political corruption, defined as the unlawful use of public office for private gain², can be an important obstacle to both economic and social development³. Corruption reduces levels of international and national investment, lowers government spending on public services and favours the establishment of projects that allow the misappropriation of funds⁴. This has led to demands by some donor agencies that their future economic assistance to corrupt countries is linked to improved governance and accounting procedures. At the same time, the donor community is being pressed to increase funding to reduce the threat of biodiversity loss and to build conservation capacity¹. The sums needed to conserve biodiversity are small relative to its total economic value^{1,5,6}, but still large in absolute terms⁷. Therefore, current debates centre on ensuring that conservation funds are raised and spent efficiently^{8,9}.

Although evidence from case studies suggests that corruption may affect the outcomes of conservation projects^{10–13}, no study has formally assessed the broader significance of governance to international conservation efforts. This is surprising, because there are several reasons for expecting conservation to be particularly susceptible to subversion by corrupt officials. First, many projects are short term and funded externally, increasing the scope for misappropriation of funds. Second, most government officials are poorly paid, which encourages the acceptance of bribes, and this is particularly likely when they are responsible for managing natural resources with high financial value. Third, conservation departments tend to lack political weight, making it difficult to prevent transgressions by other sectors, such as the armed forces¹³. Finally, measuring the success of conservation projects is difficult¹⁴, allowing corrupt practices to remain hidden. Based on these concerns, we sought to investigate the potential effect of corruption on biodiversity conservation by addressing two key questions. First, are changes in three well surveyed and widespread components of biodiversity—namely forests, African elephants and black rhinoceroses—correlated with governance scores? Second, is there any link between those countries that have been identified as conservation priorities and governance scores?

National governance scores were based on a system that assigns a maximum value of 10 to the least-corrupt countries². Such scores may also be related to a number of other socio-economic factors, and so a preliminary analysis tested for correlations between governance scores and per capita gross domestic product (GDP), Human Development Index (HDI) scores and population density (see Methods). Using Pearson's (r) and Spearman's (r_s) correlations, governance scores were found to be positively correlated with per capita GDP ($n = 75$, $r = 0.847$, $P < 0.001$) and HDI scores ($n = 73$, $r = 0.713$, $P < 0.001$), but not with population density ($n = 77$, $r_s = 0.045$, $P = 0.697$), and all four factors were used in subsequent correlation and stepwise multiple regression analyses.

Our first main analysis investigated links between national governance scores and changes in forest cover. This was because illegal logging is a problem in many countries, and data on forest cover change are available for a large number of nations. Previous studies have related these changes to poverty levels, population density and governance types^{15,16}. Therefore, we tested whether national changes in forest cover between 1990 and 1995 were associated with mean governance scores, as well as mean per capita GDP, mean HDI scores and mean population density (see Sup-

plementary Information). We found that change in total forest cover, which included plantations, was correlated with mean per capita GDP ($n = 101$, $r_s = 0.577$, $P < 0.001$) and mean governance scores ($n = 106$, $r_s = 0.556$, $P < 0.001$; Fig. 1). However, change in natural forest was not correlated with mean governance scores ($n = 83$, $r_s = 0.181$, $P = 0.102$), suggesting that the result for total cover was driven by the establishment of new plantations in wealthier, better-governed countries. This might seem surprising, as corruption has been linked to deforestation¹⁷. However, logging is also used more widely as an economic development policy, which may explain why no pattern was observed.

Corruption might play a larger role when trying to circumvent trade bans underpinned by international law. The African elephant and black rhinoceros are now fully protected by the Convention on International Trade in Endangered Species (CITES), in response to continued demand for ivory and rhinoceros horn. A previous study related population changes in the early 1980s to national spending per km² of protected area (PA)¹⁸. Therefore, we used stepwise multiple regression modelling for which we give both degrees of freedom ($f_{x,y}$) and the regression coefficient (B), to find whether mean governance scores, national spending per km² of PA, mean per capita GDP, mean HDI scores and mean human population densities were related to changes in elephant and rhinoceros numbers between 1987 and 1994^{19,20}, a period when numbers of both species declined (see Supplementary Information). We found that changes in numbers during this period were related only to governance score, for both African elephants ($F_{1,20} = 20.38$, $B = 27.37$, $P < 0.001$, $r^2 = 0.531$; Fig. 2a) and black rhinoceroses ($F_{1,9} = 20.33$, $B = 36.93$, $P = 0.003$, $r^2 = 0.744$; Fig. 2b).

These results suggest that political corruption may play a considerable role in determining the success of national strategies to conserve these two flagship species, despite the international attention they both attract. National spending per km² of PA was not related to population change for either species (African elephants: $P = 0.218$; black rhinoceroses: $P = 0.386$), but this change from the early 1980s probably reflects the development of conservation strategies that focus greater financial resources on protecting key populations, the effectiveness of which is not captured in data on national spending. Such strategies have helped numbers of both species to increase since 1994, but stepwise multiple regression modelling showed that the strength of both recoveries between 1994 and 1998 was again related to mean governance scores alone (African elephants using log₁₀ modelled governance scores:

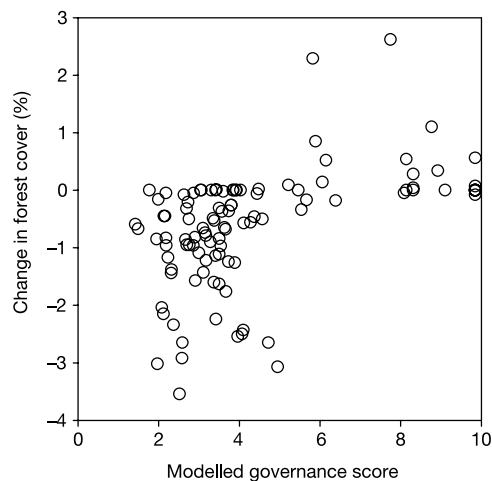


Figure 1 Mean modelled governance scores and changes in total national forest cover, 1990–95.

$F_{1,18} = 7.06$, $B = 181.55$, $P = 0.017$, $r^2 = 0.306$; black rhinoceroses: $F_{1,9} = 19.63$, $B = 32.23$, $P = 0.003$, $r^2 = 0.737$). The model for African elephants during this recovery period predicts that countries with governance scores of less than 3.1 would show population declines. Interestingly, the mean governance score in 2002 for the six Asian elephant range states with available data is 2.9, and total Asian elephant numbers have continued to decline, despite the CITES ban on trade in ivory.

As results suggest that corruption may affect the success of conservation efforts, we also investigated corruption levels in countries identified as global priorities for conservation funding (see Supplementary Information). There is indeed a negative relationship between governance scores and combined bird and mammal species richness, adjusted for country area ($n = 83$, $r_s = -0.473$, $P < 0.001$; Fig. 3). However, most non-governmental organizations (NGOs) that donate funds for conservation do not set their national priorities on species richness alone. Therefore, we investigated the governance scores of countries that contain priority areas identified by three NGOs, based on endemism, threat and representativeness. We found that countries containing Endemic Bird Areas²¹ (EBAs), biodiversity hotspots²² and Focal 25 terrestrial ecoregions²³ all had lower governance scores than those countries that did not (Table 1).

These results highlight the need for the conservation community to develop approaches that offset the effects of poor governance.

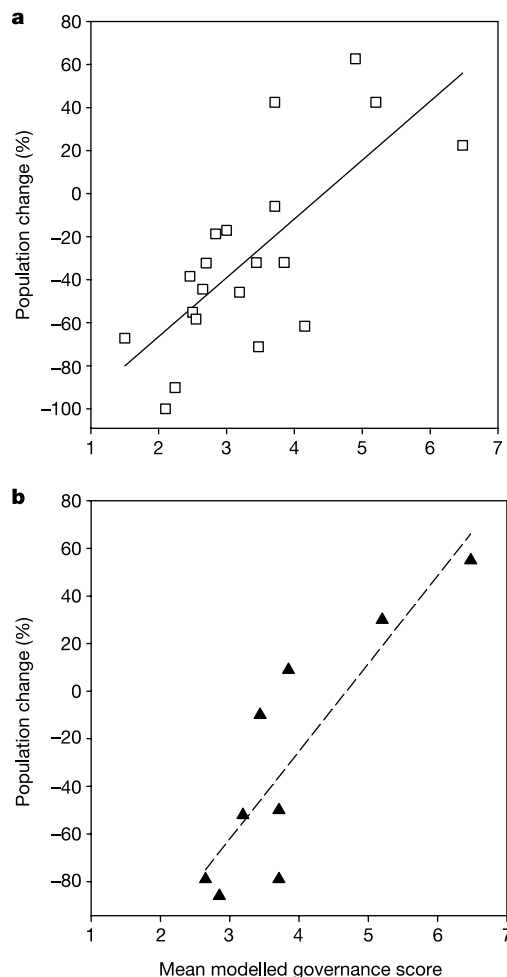


Figure 2 Mean modelled governance scores and changes in national populations of two species, 1987–94. **a**, African elephants; **b**, black rhinoceroses.

Table 1 Governance scores and priority areas for conservation

Priority area type	Mean governance scores		t statistic
	Contains	Outside	
BirdLife EBAs	3.65	6.37	5.61*
CI hotspots	3.86	5.86	3.90*
WWF Focal 25	3.31	5.42	4.62*

EBA, Endemic Bird Area; CI, Conservation International; WWF, WorldWide Fund for Nature.

* $P < 0.001$

They also support the ample anecdotal evidence that links corruption and conservation failure both within PAs, where staff can be bribed to allow transgressions, and outside PAs, where communities lack the capacity to protect natural resources from corrupt officials²⁴. Simplistically, our results might suggest focusing donor investment on better governed countries. However, this would be unwise, as many countries with poor governance have high levels of species endemism, as well as distinctive populations of widespread species, and should not be neglected. Instead, successful conservation projects in poorly governed countries show that these problems can be reduced by developing a motivated, well paid workforce²⁵, using more stringent accounting procedures and engaging the private sector in management partnerships²⁶.

The international donor community should, therefore, continue to use their influence to encourage appropriate reforms²⁷. They should also publicly differentiate between genuine conservation priorities and crisis events caused by poor governance, avoiding the temptation to use the latter as fund-raising opportunities or reasons for changing international policies. However, there is also a need for a broad review of present conservation practice to develop systems that are less open to subversion. For example, outlawing international trade in endangered species continues to be championed as a universally applicable approach to conservation. This is despite evidence that these bans are ineffective in poorly governed countries if demand remains high, and that they encourage bribery and increase the power of corrupt officials²⁸. At present, any policy review is hampered by a lack of relevant data and an unwillingness to discuss project-based issues associated with corruption. Instead, the conservation community needs to accept that these problems must be identified and discussed openly if they are to be resolved. □

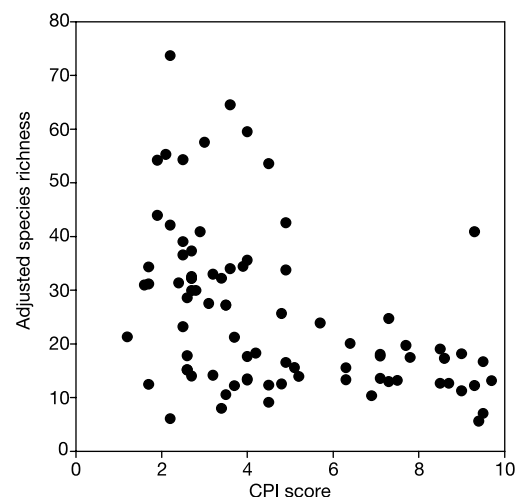


Figure 3 National governance scores and adjusted species richness.

Methods

The governance scores were based on the Corruption Perception Index (CPI) produced by Transparency International². This system uses independent surveys of business people and assessments by country analysts to compare national corruption levels. However, it was only initiated in 1995, so another data set, the International Country Risk Guide (ICRG), was used to provide information for earlier years. The ICRG system uses several coarse-scale factors to measure national governance levels, and a data set was available that contained information on 126 countries between 1984 and 1999. Therefore, we used stepwise linear regression analysis to produce a model that allowed log₁₀ CPI scores to be calculated on the basis of these ICRG data. This model was developed using data for 1999 and included three ICRG factors, which were 'corruption' ($B = 0.074$, $P < 0.001$), 'bureaucratic quality' ($B = 0.0057$, $P < 0.001$) and 'law and order' ($B = 0.0028$, $P = 0.008$). In this system, 'corruption' measures corruption within the political system, 'bureaucratic quality' measures the ability of the bureaucracy to govern without drastic changes in policy or implementation, and 'law and order' measures both popular observance of the law and the strength and impartiality of the legal system. The model explained most of the observed variation ($F_{3,91} = 122.68$, adjusted $r^2 = 0.809$) and comparisons of predicted and actual CPI scores for 1995 and 1996 showed that the model had high levels of explanatory power, which were increased in the latter period when data on a more representative number of developing countries were available (1995: $n = 40$, $r^2 = 0.728$; 1996: $n = 53$, $r^2 = 0.860$). Therefore, we used this model to calculate CPI scores for those analyses that used biodiversity data collected before 1995.

Data obtained by the Food and Agriculture Organization on changes in forest cover between 1990 and 1995, together with information on African elephant populations¹⁹ and black rhinoceros populations²⁰ were collated as the best available data on changes in widespread biodiversity elements. We also collated data on annual per capita Gross Domestic Product (GDP), Human Development Index (HDI) scores produced by the United Nations Development Programme and human population density data. Data on national conservation budgets were available from one of a range of years from 1991 to 1996 for a number of African countries²⁹, so these were adjusted to 1993 US\$ (the median date) using deflation indexes produced by the International Monetary Fund. Spearman's rank correlation tests and stepwise multiple regression modelling were then used to identify factors that were related to national percentage change in these biodiversity components. These factors were transformed, whenever necessary, to meet the assumptions of the tests. In each case, data from countries with a restricted amount of each component were excluded, as small changes or measurement errors were more likely to produce apparently extreme results. The exclusion levels were: forest area <30,000 km², elephant populations <1,000 and black rhinoceros populations <10.

We used CPI scores for 2002 to investigate the relationships between national biodiversity levels and governance. Species richness values for each country were calculated as the number of recorded mammal and bird species, using data available from the World Resource Institute. We corrected for the nonlinear relationship between species number and country area by dividing species number by A^z , where A is country area, and z is a typical value for the slope of a nested, within-continent plot of log(species number) on log(area), set here as 0.25 (ref. 30).

Received 6 June; accepted 22 August 2003; doi:10.1038/nature02025.

- Balmford, A. *et al.* Ecology—Economic reasons for conserving wild nature. *Science* **297**, 950–953 (2002).
- Transparency International *Corruption Perceptions Index 2002* (Transparency International, Berlin, 2002).
- Kaufmann, D. Corruption: The facts. *Foreign Policy* **107**, 114–131 (1997).
- Azfar, O., Lee, Y. & Swamy, A. The causes and consequences of corruption. *Ann. Am. Acad. Polit. Sci.* **573**, 42–56 (2001).
- Costanza, R. *et al.* The value of the world's ecosystem services and natural capital. *Nature* **387**, 253–260 (1997).
- Balmford, A., Gaston, K. J., Blyth, S., James, A. & Kapos, V. Global variation in terrestrial conservation costs, conservation benefits, and unmet conservation needs. *Proc. Natl Acad. Sci. USA* **100**, 1046–1050 (2003).
- James, A., Gaston, K. J. & Balmford, A. Can we afford to conserve biodiversity? *Bioscience* **51**, 43–52 (2001).
- Hughey, K. F. D., Cullen, R. & Moran, E. Integrating economics into priority setting and evaluation in conservation management. *Conserv. Biol.* **17**, 93–103 (2003).
- Balmford, A. & Whitten, T. Who should pay for tropical conservation, and how could the costs be met? *Oryx* **37**, 238–250 (2003).
- Barnes, R. F. W., Blom, A. & Alers, M. P. T. A review of the status of forest elephants *Loxodonta africana* in Central Africa. *Biol. Conserv.* **71**, 125–132 (1995).
- Archabal, K. & Naughton-Treves, L. Tourism revenue-sharing around national parks in Western Uganda: Early efforts to identify and reward local communities. *Environ. Conserv.* **28**, 135–149 (2001).
- Myers, N. *The Primary Source: Tropical Forests and our Future* (Norton & Co, New York, 1992).
- Robertson, J. M. Y. & van Schaik, C. P. Causal factors underlying the dramatic decline of the Sumatran orang-utan. *Oryx* **35**, 26–38 (2001).
- Salafsky, N., Margoluis, R., Redford, K. H. & Robinson, J. G. Improving the practice of conservation: A conceptual framework and research agenda for conservation science. *Conserv. Biol.* **16**, 1469–1479 (2002).
- Didia, D. O. Democracy, political instability and tropical deforestation. *Glob. Environ. Change* **7**, 63–76 (1997).
- Geist, H. J. & Lambin, E. F. Proximate causes and underlying driving forces of tropical deforestation. *Bioscience* **52**, 143–150 (2002).
- Jepson, P., Jarvie, J. K., MacKinnon, K. & Monk, K. A. The end for Indonesia's lowland forests? *Science* **292**, 859 (2001).
- Leader-Williams, N. & Albon, S. D. Allocation of resources for conservation. *Nature* **336**, 533–535 (1988).

- Barnes, R. F. W. *et al.* *African Elephant Database 1998* (IUCN/SSC African Elephant Specialist Group, IUCN, Gland, Switzerland, 1999).
- Emslie, R. & Brooks, M. *African Rhino: Status Survey and Conservation Action Plan* (IUCN/SSC African Rhino Specialist Group, IUCN, Gland, Switzerland, 1999).
- Stattersfield, A. J., Crosby, M. J., Long, A. J. & Wege, D. C. *Endemic Bird Areas of the World: Priorities for Biodiversity Conservation* (BirdLife International, Cambridge, UK, 1998).
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
- World Wide Fund For Nature *Living Planet Campaign: A Call to Action* (WWF, Washington DC, 1999).
- Whitten, T., Holmes, D. & MacKinnon, K. Conservation biology: A displacement behavior for academia? *Conserv. Biol.* **15**, 1–3 (2001).
- Leakey, R. & Morell, V. *Wildlife Wars: My Battle to Save Kenya's Elephants* (Pan Macmillan, London, 2001).
- Walpole, M. J. & Leader-Williams, N. Masai Mara tourism reveals partnership benefits. *Nature* **413**, 771 (2001).
- International Institute for Environment and Development *Whose Eden? An Overview of Community Approaches to Wildlife Management* (IIED, London, 1994).
- Oldfield, S. *The Trade in Wildlife: Regulation for Conservation* (Earthscan, London, 2002).
- James, A., Green, M. J. B. & Paine, J. R. (eds) *A Global Review of Protected Area Budgets and Staff* (WCMC-World Conservation Press, Cambridge, UK, 1999).
- Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, UK, 1995).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the Durrell Institute for Conservation and Ecology for providing funding, and K. Brandon, N. Burgess, S. Elton, R. Green, J. Groombridge, A. Huggins, S. Thornton and T. Whitten for advice.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.J.S. (R.J.Smith@kent.ac.uk).

Sophisticated sperm allocation in male fowl

Tommaso Pizzari^{1*}, Charles K. Cornwallis², Hanne Løvlie³, Sven Jakobsson³ & Tim R. Birkhead²

¹Department of Animal Environment & Health, Swedish University of Agricultural Sciences, Skara PO Box 234, SE-532 23, Sweden

²Department of Animal & Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

³Department of Zoology, Stockholm University, Stockholm SE-106 91, Sweden

* Present address: School of Biology, The University of Leeds, Leeds LS2 9JT, UK

When a female is sexually promiscuous, the ejaculates of different males compete for the fertilization of her eggs¹; the more sperm a male inseminates into a female, the more likely he is to fertilize her eggs². Because sperm production is limited and costly, theory predicts that males will strategically allocate sperm (1) according to female promiscuity^{1,3–5}, (2) saving some for copulations with new females^{3,6,7}, and (3) to females producing more and/or better offspring^{3,8}. Whether males allocate sperm in all of these ways is not known, particularly in birds where the collection of natural ejaculates only recently became possible. Here we demonstrate male sperm allocation of unprecedented sophistication in the fowl *Gallus gallus*. Males show status-dependent sperm investment in females according to the level of female promiscuity; they progressively reduce sperm investment in a particular female but, on encountering a new female, instantaneously increase their sperm investment; and they preferentially allocate sperm to females with large sexual ornaments signalling superior maternal investment. Our results indicate that female promiscuity leads to the evolution of sophisticated male sexual behaviour.

In the fowl, socially dominant males have privileged copulatory

access to females and females copulate with several males in each breeding cycle^{9,10}. The resulting sperm competition leads males to allocate sperm differentially to females according to female promiscuity, female novelty and female reproductive quality.

We found that males differentially allocated sperm according to female promiscuity, which was simulated by experimentally exposing a male to a female in the presence of zero, one or three male competitors, and according to the male's own social status. In the absence of competitors, males minimized their sperm investment. As the number of competitors increased from one to three, dominant males increased their sperm investment, but subdominants maximized their investment in the presence of just one competitor (Fig. 1).

A male's propensity to copulate with a particular female declined with the time that a male was exposed to a female, but was renewed by replacing the female with a new female (Fig. 2). After an initial bout of frequent copulations immediately after exposure to a female, males left to forage and returned regularly to inspect the female's head closely, suggesting that males may use visual cues to recognize individual females when making copulation decisions. Once a male lost interest in a female, the female was exchanged for another and the male resumed copulating. When we removed and re-presented the same female after a male had ceased copulating with her, however, the male inspected the female but in no case copulated with her ($n = 5$ trials).

Males progressively reduced their sperm investment in a particular female according to the number of sperm that they had already inseminated into her, but increased their investment when allowed to copulate with a new female (Fig. 3a). Successive females obtained fewer sperm, confirming that male sperm reserves were depleted over successive copulations. But although males allocated progressively fewer sperm to successive ejaculations with an individual female, they increased their sperm investment when presented with a new female, indicating preferential sperm investment in new females.

In several bird species including the fowl, some copulations do

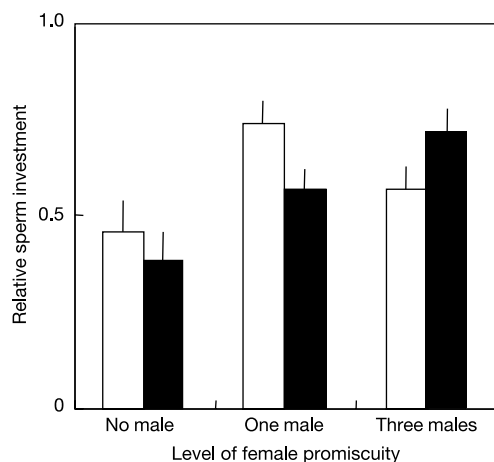


Figure 1 Differential sperm allocation and female promiscuity. Relative sperm investment of dominant (filled) and subdominant (open) male feral fowl according to female promiscuity: dominants steadily increased relative sperm investment as the number of competitors (audience) increased, whereas subdominants decreased their investment when the number of competitors increased beyond one. We measured relative sperm investment as the cumulative number of sperm transferred by a male during a trial standardized against his largest cumulative number of sperm produced in a trial across all audience treatments (generalized linear mixed model with restricted maximum likelihood estimation¹⁶ (REML–GLMM) of relative sperm investment, with male status and audience as fixed effects, and male identity as a random effect: status, $F_{1,34} = 0.11$, $P = 0.75$; audience, $F_{1,34} = 3.90$, $P = 0.030$; status $\times$ audience, $F_{2,34} = 4.20$, $P = 0.024$). Bars indicate s.e.m.

not result in the transfer of semen^{11–14}. Consistent with the idea that male fowl retain some of their sperm reserves for other females, the probability of a male failing to transfer sperm increased significantly over successive copulations with a particular female, regardless of the number of females to which a male had been exposed (Fig. 3b). We also replicated this experiment with a population of red jungle fowl, *G. gallus*, the wild ancestor of the domestic fowl¹⁵, and obtained similar results (Figs 2 and 3, legend).

Males preferentially invested sperm in females with relatively large sexual ornaments. When experimentally exposed to two females simultaneously, males were more likely to copulate with (Fig. 4a) and allocate more sperm to (Fig. 4b) the female with the larger comb. Sperm allocation was also status-specific: socially dominant males biased sperm investment in favour of the large-combed female more than did subdominant males (Fig. 4b).

We investigated the adaptive significance of male preference for large-combed females. First, focal watches of free-ranging groups of four females and two males ($n = 7$) showed that females with larger combs (ranked according to comb size in a group) received significantly more successful copulations from both males (generalized linear model with restricted maximum likelihood estimation (REML–GLM)¹⁶ of number of copulations, with comb rank and male status as fixed effects nested in groups, and male identity as random effect: comb rank, $X^2_{48} = 37.17$, $P = 0.023$; male status, $X^2_{48} = 8.07$, $P = 0.018$; male status $\times$ comb rank, $X^2_{48} = 24.19$, $P = 0.062$). Second, female comb size was an important predictor of mean egg mass (Fig. 4c) and mean yolk mass of eggs, after controlling for mean egg mass (yolk mass = $1.07 \pm 0.892 + 0.43 \pm 0.077$ (egg mass) + 0.002 ± 0.001 (comb mass) – 0.002 ± 0.001 (body mass), $R^2 = 0.72 \pm 0.201$, $F_{3,25} = 18.5$, $P < 0.0001$; comb mass, $t = 2.13$, $n = 26$, $P = 0.045$). These results indicate that large-combed females are subject to higher sperm competition, produce larger eggs and allocate more resources to embryos.

Our results indicate that sexual promiscuity combined with status-specific constraints causes high behavioural flexibility in male sperm allocation. Dominant and subdominant male fowl face different levels of sperm competition, and tailor their ejaculates accordingly within the constraints imposed by their status. Theory

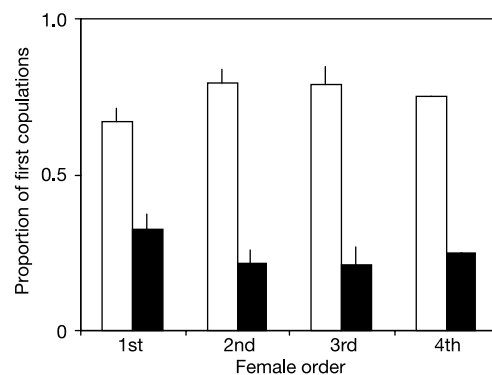


Figure 2 Preferential sperm investment in new females. In feral fowl, male propensity to copulate declined with the time that a male was exposed to a female, and increased with exposure to a new female: males copulated more frequently during the first (open) than the second (filled) half of the time that they were exposed to a female, regardless of with how many females they had previously copulated (REML–GLMM of number of copulations, with female order and first and second halves of the exposure time to a female, male identity and year as fixed effects, and duration of exposure time (min) as a covariate: female order, $F_{3,144} = 0.30$, $P = 0.83$; half, $F_{1,147} = 56.09$, $P < 0.0001$; duration, $F_{1,147} = 123.79$, $P < 0.0001$; order $\times$ half, $F_{1,143} = 1.81$, $P = 0.15$). Bars represent s.e.m. A similar pattern was seen in jungle fowl (REML–GLM; female order, $X^2_{2,39} = 3.65$, $P = 0.16$; half, $X^2_{1,41} = 3.95$, $P = 0.04$; duration, $X^2_{1,41} = 57.75$, $P < 0.0001$; order $\times$ half, $X^2_{2,37} = 0.55$, $P = 0.76$).

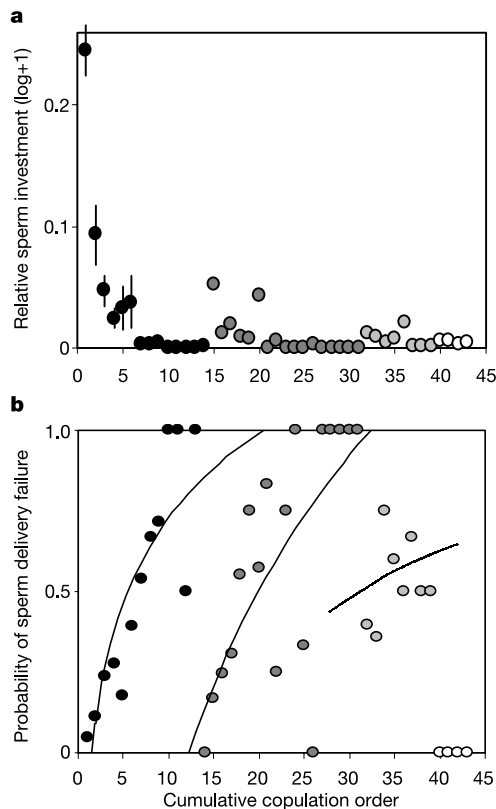


Figure 3 Preferential sperm investment in new females (continued). **a**, Relative sperm investment of male feral fowl declined over successive inseminations with the same female and was renewed by the presence of a new female. Relative sperm investment during a trial was measured as the number of sperm in an ejaculate standardized against the largest ejaculate that a male produced across all females in the same trial (REML–GLMM of relative sperm investment, with female order as a fixed effect, sexual familiarity measured as the order of copulation of a male within each female as a covariate, and year and male identity as random effects: female order, $F_{3,367} = 10.57$, $P < 0.0001$; sexual familiarity, $F_{1,367} = 38.40$, $P < 0.0001$; order $\times$ familiarity, $F_{3,364} = 3.31$, $P = 0.76$). Bars represent s.e.m. A similar pattern was seen in male jungle fowl (REML–GLM with female order as a fixed effect, and sexual familiarity as a covariate: female order, $X^2_{2,163} = 23.80$, $P < 0.0001$; sexual familiarity, $X^2_{1,163} = 12.61$, $P = 0.0004$; order $\times$ familiarity, $X^2_{2,161} = 0.03$, $P = 0.986$). Consistent with differential sperm allocation, sperm investment in the first ejaculate with a female was not lower than that in the last ejaculate with the previous female. Males tended to invest more sperm in the first ejaculate with a new female than in the last ejaculate with the familiar female (Wilcoxon paired tests: number of sperm in first ejaculate with second female versus last ejaculate with first female: feral fowl, $Z = -2.29$, $P = 0.02$, sign test $P = 0.057$, $n = 17$; jungle fowl, $Z = -0.16$, $P = 0.116$, sign test $P = 0.69$, $n = 7$. First ejaculate with third female versus last ejaculate with second female: feral fowl, $Z = -0.52$, $P = 0.60$, sign test $P = 0.69$, $n = 6$; jungle fowl, $Z = -1.83$, $P = 0.07$, sign test $P = 0.125$, $n = 6$). **b**, Mean probability of semen delivery failure over successive copulations in male feral fowl. Males were more likely to copulate without ejaculating semen the more they copulated with a female (REML–GLM of mean probability of failure of semen transfer, with female order, male identity and year as fixed effects, and sexual familiarity as a covariate: female order, $X^2_{2,292} = 15.91$, $P = 0.0012$; sexual familiarity, $X^2_{1,292} = 59.39$, $P < 0.0001$. Similar results were found in male jungle fowl (female order, $X^2_{2,163} = 16.91$, $P < 0.0002$; sexual familiarity, $X^2_{1,163} = 14.72$, $P < 0.0001$). In **a**, **b**, black symbols indicate first female; dark grey, second; light grey, third; open, fourth.

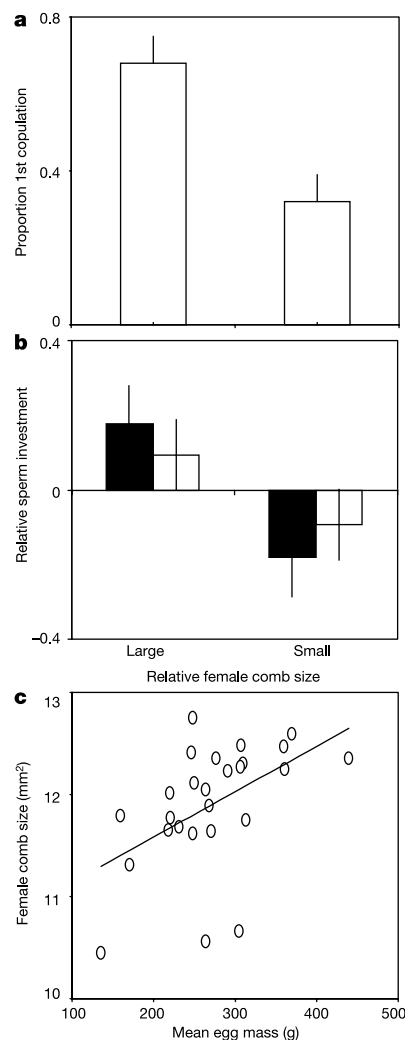


Figure 4 Mate choice and strategic sperm allocation by male fowl. **a**, When presented with two females simultaneously, male feral fowl were significantly more likely to mount the female with a relatively large comb. In each trial, females were randomly designated as A or B and their relative comb size was ranked as either large or small. The probability that female A was mounted first by a male was analysed in relation to whether she had a large or small comb (REML–GLM of binary measure of whether female A was mounted first, with female comb rank and male status as fixed effects, male identity and male groups as random effects, and female relative body mass and body size (A – B) as covariates: female comb rank, $F_{2,53} = 8.96$, $P = 0.0042$). This preference was independent of male social status (male status $\times$ female comb rank nested within male pairs: $F_{2,48} = 0.35$, $P = 0.71$; all other effects not significant). **b**, Male feral fowl allocated significantly more sperm to females with relatively large combs (REML–GLMM of relative sperm investment measured as the number of sperm invested in female A minus that invested in female B in each trial, with copulation order to control for the effect of sperm depletion, female comb rank and male status as fixed effects, male identity and male group as random effects, and relative female body mass and body size as covariates: copulation order, $F_{1,119} = 57.23$, $P < 0.0001$; female comb rank, $F_{1,119} = 11.58$, $P = 0.0009$), but dominant males (filled) invested more sperm than did subordinate males (open) in large-combed females (status $\times$ comb rank nested within male pairs: $F_{2,119} = 3.91$, $P = 0.02$; all other effects not significant). **c**, Comb size, x , was a predictor of the average egg mass produced by a female feral fowl, y ($y = 10.7 \pm 0.45 + 0.004 \pm 0.002x$; $R^2 = 0.24 \pm 0.54$, $F_{1,25} = 7.75$, $P < 0.01$; comb size: $t = 2.78$, $n = 26$, $P < 0.01$). Removing the three lowest data points strengthens the regression ($R^2 = 0.357 \pm 0.31$, $P < 0.003$).

predicts that males increase their sperm investment in a female with increasing risk of sperm competition (that is, the competition for fertilization between the ejaculates of different males generated by female promiscuity¹). When females obtain sperm from several males however, sperm competition is certain and males are predicted to maximize sperm investment in those females that copulate with just one other male, and to decrease sperm investment progressively as the number of males that inseminate a female increases (increasing sperm competition intensity^{1,3}).

The response of dominant males to different levels of sperm competition is consistent with the strategy predicted under risk of sperm competition^{1,3}, and the sperm allocation by subordinate males with that predicted under sperm competition intensity^{1,3}. Dominant males have privileged control over copulations, and the presence of other males may represent an increased risk of another male inseminating the same female (but not necessarily an increased number of males inseminating this female). As subordinate males cannot prevent females from copulating with other males¹⁰, the presence of other males is likely to result in more intense sperm competition. Although copulation success co-varies with male status in the fowl, most subordinate males still experience multiple copulations⁹, and therefore may be selected to save sperm for less-competitive copulation opportunities. In addition, status may be condition-dependent in male fowl¹⁷, and thus sperm investment may be more costly for subordinate males.

The decline in male sexual interest in the same female and its resuscitation with a new female is known as the Coolidge effect¹⁸. The Coolidge effect has been thought to mediate differential sperm allocation, enabling a male to distribute sperm more evenly and adaptively across multiple females^{3,19}. Because the number of sperm inseminated by a male into different females has been seldom quantified, the adaptive significance of the Coolidge effect was previously unknown. Our study shows experimentally that a male's propensity to copulate is matched by the number of sperm transferred, thus indicating that the Coolidge effect may be adaptive to males.

Our study also shows that under sexual promiscuity male choice of partners occurs at two levels: behaviourally and cryptically through differential sperm allocation. Although the male comb is known to be important in partner choice in the fowl²⁰, the function of the female comb has received less attention. Male preference for more ornamented partners is explained by the superior condition and reproductive investment of these females. Prudence in sperm investment and preferential sperm allocation to more ornamented females by dominant males are consistent with dominant males having privileged access to females¹⁰, and indicate that dominant males may be more competitive when they inseminate high-quality females, suggesting an important but neglected reason for why high-quality males may produce high-quality offspring. □

Methods

Study populations

We studied a population of feral fowl (11–16 males, 28–32 females), free-ranging at the Tovetorp Zoological Research Station, Sweden^{9,10,21,22}, in June–July 2000, April–September 2001 and April–June 2002. The female order experiment was also replicated in a red jungle fowl population²¹ (16 males, 30 females) at the same site (June–July 2000). Males were kept isolated from females and sexually rested for at least 2 d to allow them to replenish their sperm reserves¹¹. We collected natural ejaculates by presenting males with a live female fitted with a harness for collecting the ejaculate and held in a soliciting position to minimize the possibility that female behaviour would influence male copulatory response^{9,23}. All males were fully habituated to human presence and no alteration in the birds' normal behaviour was observed during the trials. After copulation, ejaculates were always collected in the same manner⁹ and the sperm were counted²⁴. Ejaculates obtained with this technique are similar in volume to ejaculates naturally inseminated by males and subsequently ejected by females⁹.

Female promiscuity experiment

We collected natural ejaculates of male feral fowl allowed to copulate *ad libitum* with a female in the presence of three, one and no other males. Males were kept in pens in groups of four, two and single males: 16 males were exposed to the four-male treatment (four

groups with two top- and two bottom-ranking²⁵ males), and were rearranged into eight pairs of one top- and one bottom-ranking male (the top- with the third-ranking, and the second- with the fourth-ranking) in the two-male treatment. Eight of these sixteen males (four top- and four bottom-ranking males, randomly chosen) were also exposed individually to the one-male treatment. We randomized the order of treatments across males. Males were allowed singly out of the pen to copulate with a female less than 5 m from the pen in full view of the other group members.

Female novelty experiment

Each male was isolated 2 h before each trial and then allowed to copulate *ad libitum* with a female until he had lost sexual interest in her (at least 10 min had elapsed since he last ejaculated), after which the female was replaced by a new one. The new female was kept out of sight of the focal male until the familiar female was removed so that a male was never exposed to more than one female at any given time. In 2000, we used ten male feral fowl, which were each replicated (mean $\pm$ s.e.m.) 2.20 ± 0.29 (with first female), 1.78 ± 0.28 (first and second) and 2.00 ± 0.41 (first, second and third) times. In 2002, we replicated the experiment on 12 different male feral fowl, which were each exposed to one trial.

All 22 male feral fowl copulated and ejaculated sperm when exposed to a first female; 17 (77%) ejaculated sperm when presented with a second female, and 6 of these 17 (53%) ejaculated sperm with a third female. These six males also copulated with a fourth female, and one of them (17%) ejaculated sperm with this female (Figs 2 and 3). Male jungle fowl were exposed to one trial each: seven of nine (78%) ejaculated sperm when presented with a second female, six of these seven (86%) also copulated with a third female.

Female reproductive quality experiment

Eight male pairs were kept in pens in visual (but not physical) contact with a group of four females. Male hierarchies were assessed²⁵. The focal male was isolated 2 h before the trial and then presented with two females from the group simultaneously for 1 min, after which he was allowed to copulate once with one of them. Copulation with the other female was then encouraged by placing a wire cage over the first female. If copulation with a female did not occur within 15 min, sperm investment in that female was considered zero. After males were sexually rested, the trial was repeated but the order of copulation with each female was reversed (thus, a male could copulate with each female on two occasions, 48 h apart).

Each male was replicated using four different groups of females: with two groups, males were allowed to choose their first copulation partner; with the remaining two, the copulation order was predetermined. Thus, male mate choice was assessed by two replicates and sperm allocation by all four replicates. Female comb area was calculated with Photoshop (Adobe) from a digital image in standard light conditions and females were ranked according to their relative comb size in all trials.

Eggs were collected for 51 d (12 May to 2 July 2002), dissected into the constituent parts, baked at 40 °C for 12 h, and dry-weighted to 0.0001 g. Maternity was established through yolk staining with lipid dyes fed to females. We measured female body size using PC1 of principle component analysis of head, tarsus and wing length (79.2% of variation explained).

Received 3 June; accepted 16 August 2003; doi:10.1038/nature02004.

- Parker, G. A. in *Sperm Competition and Sexual Selection* (eds Birkhead, T. R. & Møller, A. P.) 3–54 (Academic, London, 1998).
- Martin, P. A., Reimers, T. J., Lodge, J. R. & Dziuk, P. J. The effect of ratios and numbers of spermatozoa mixed from two males on proportions of offspring. *J. Reprod. Fertil.* **39**, 251–258 (1974).
- Wedell, N., Gage, M. J. W. & Parker, G. A. Sperm competition, male prudence and sperm-limited females. *Trends Ecol. Evol.* **17**, 313–320 (2002).
- Pilastro, A., Scaggiante, M. & Rasotto, M. B. Individual adjustment of sperm expenditure accords with sperm competition theory. *Proc. Natl Acad. Sci. USA* **99**, 9913–9915 (2002).
- Evans, J. P., Pierotti, M. & Pilastro, A. Male mating behavior and ejaculate expenditure under sperm competition risk in the eastern mosquitofish. *Behav. Ecol.* **14**, 268–273 (2003).
- Dewsbury, D. A. Ejaculate cost and male choice. *Am. Nat.* **119**, 601–610 (1982).
- Pitnick, S. & Markow, T. A. Male gametic strategies: sperm size, testes size, and the allocation of ejaculate among successive mates by the sperm-limited fly *Drosophila pachea* and its relatives. *Am. Nat.* **143**, 785–819 (1994).
- Reinhold, K., Kurtz, J. & Engqvist, L. Cryptic male choice: sperm allocation strategies when female quality varies. *J. Evol. Biol.* **15**, 201–209 (2002).
- Pizzari, T. & Birkhead, T. R. Female feral fowl eject sperm of subordinate males. *Nature* **405**, 787 (2000).
- Pizzari, T. Indirect partner choice through manipulation of male behaviour by female fowl, *Gallus g. domesticus*. *Proc. R. Soc. Lond. B* **268**, 181–186 (2001).
- Parker, J. E., McKenzie, F. F. & Kempster, H. L. Fertility in the domestic fowl. *Miss. Agric. Exp. Station Res. Bull.* **347**, 3–50 (1942).
- Birkhead, T. R., Pellatt, J. E. & Hunter, F. M. Extra-pair copulation and sperm competition in the zebra finch. *Nature* **334**, 60–62 (1988).
- Westneat, D. F., McGraw, L. A., Fraterrigo, J. M., Birkhead, T. R. & Fletcher, F. Patterns of courtship behavior and ejaculate characteristics in male red-winged blackbirds. *Behav. Ecol. Sociobiol.* **43**, 161–171 (1988).
- Hunter, F. M., Harcourt, R., Wright, M. & Davis, L. S. Strategic allocation of ejaculates by male Adélie penguins. *Proc. R. Soc. Lond. B* **267**, 1541–1545 (2000).
- Fumihito, A. et al. Monophyletic origin and unique dispersal patterns of domestic fowls. *Proc. Natl Acad. Sci. USA* **93**, 6792–6795 (1996).
- Der, G. & Everitt, B. S. *A Handbook of Statistical Analyses using SAS* 2nd edn (Chapman & Hall, Boca Raton, FL, 2001).
- Pizzari, T. Food, vigilance and sperm: the role of male direct benefits in the evolution of female preference in a polygamous bird. *Behav. Ecol.* (in the press).
- Wilson, J. R., Kuehn, R. E. & Beach, F. A. Modification in the sexual behavior of male rats produced by changing the stimulus female. *J. Comp. Physiol. Psychol.* **56**, 636–644 (1963).
- Pizzari, T. Sperm allocation, the Coolidge effect and female polyandry. *Trends Ecol. Evol.* **17**, 456 (2002).

20. Zuk, M., Johnsen, T. S. & MacLarty, T. Endocrine-immune interactions, ornaments and mate choice in red jungle fowl. *Proc. R. Soc. Lond. B* **260**, 205–210 (1995).
21. Schütz, K. E. & Jensen, P. Effects of resource allocation on behavioural strategies: a comparison of red jungle fowl (*Gallus gallus*) and two domesticated breeds of poultry. *Ethology* **107**, 753–765 (2001).
22. Pizzari, T. & Birkhead, T. R. For whom does the hen cackle? The adaptive significance of the post-oviposition cackling. *Anim. Behav.* **61**, 601–607 (2001).
23. Ishikawa, H. The life duration of cock spermatozoa outside the body. *Proc. World Poultry. Fourth Congr.* 90–94 (1930).
24. Bakst, M. R. & Cecil, H. C. (eds) *Techniques for Semen Evaluation, Semen Storage, and Fertility Determination* (Poultry Science Association, Savoy, IL, 1997).
25. Clutton-Brock, T. H., Albon, S. D., Gibson, R. M. & Guinness, F. E. The logical stag. *Anim. Behav.* **27**, 211–225 (1979).

Acknowledgements We thank A. Donald for help with sperm counting, and A. MacColl for statistical advice. T.P. was supported by a scholarship from the Swedish University of Agricultural Sciences and a grant from the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS); T.R.B. by a grant from the Natural Research Council (NERC); C.K.C. by a studentship from NERC (to T.R.B.); and H.L. by a studentship from the Zoology Department of Stockholm University. T.R.B. and T.P. developed the sperm collection technique; T.P. designed the female promiscuity and female novelty experiments, and together with C.K.C. and H.L. carried them out; C.K.C. designed the female quality experiment with T.R.B. and carried it out together with H.L.; C.K.C. analysed the data with T.P.; T.P. wrote the paper; S.J. helped with field work and the provision of research facilities.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.P. (t.pizzari@leeds.ac.uk).

Truncated TrkB-T1 mediates neurotrophin-evoked calcium signalling in glia cells

Christine R. Rose, Robert Blum, Bruno Pichler, Alexandra Lepier, Karl W. Kafitz & Arthur Konnerth

Institut für Physiologie, Ludwig-Maximilians-Universität München, 80336 München, Germany

The neurotrophin receptor TrkB is essential for normal function of the mammalian brain^{1–3}. It is expressed in three splice variants. Full-length receptors (TrkB^{FL}) possess an intracellular tyrosine kinase domain and are considered as those TrkB receptors that mediate the crucial effects of brain-derived neurotrophic factor (BDNF) or neurotrophin 4/5 (NT-4/5). By contrast, truncated receptors (TrkB-T1 and TrkB-T2) lack tyrosine kinase activity and have not been reported to elicit rapid intracellular signalling⁴. Here we show that astrocytes predominately express TrkB-T1 and respond to brief application of BDNF by releasing calcium from intracellular stores. The calcium transients are insensitive to the tyrosine kinase blocker K-252a and persist in mutant mice lacking TrkB^{FL}. By contrast, neurons produce rapid BDNF-evoked signals through TrkB^{FL} and the Na_v1.9 channel^{5,6}. Expression of antisense TrkB messenger RNA strongly reduces BDNF-evoked calcium signals in glia. Thus, our results show that, unexpectedly, TrkB-T1 has a direct signalling role in mediating inositol-1,4,5-trisphosphate-dependent calcium release; in addition, they identify a previously unknown mechanism of neurotrophin action in the brain.

A short, pulse-like application of the TrkB ligand BDNF (0.73 nM, 50 ms) to cultured astrocytes evoked a Ca²⁺ wave that spread over a distance of more than 150 µm (Fig. 1a). Ca²⁺ transients recorded in single astrocytes consisted of a fast peak, which was often followed by a biphasic recovery phase (Fig. 1b, c). Applying vehicle alone or BDNF together with a BDNF scavenger (polyclonal antibodies to BDNF) was ineffective in eliciting Ca²⁺

signals (Fig. 2h). Similar astrocytic Ca²⁺ waves can be elicited by glutamate or ATP and represent a prominent form of long-range intercellular signalling in the brain^{7–10}. Notably, however, the concentrations needed for their induction by other neuroactive substances are several orders of magnitudes higher. Thus, BDNF, which is effective at subnanomolar concentrations, is the most potent endogenous agonist for the production of glia Ca²⁺ signals described so far.

With a similarly high efficiency, BDNF depolarizes neurons⁵. In contrast to the propagating Ca²⁺ wave in astrocytes, however, BDNF puffs cause neuronal Ca²⁺ signals that are spatially restricted, for example, to small portions of spiny dendrites¹¹. These neuronal BDNF-induced Ca²⁺ transients are caused by Ca²⁺ influx through voltage-gated Ca²⁺ channels after the TrkB^{FL}-mediated depolarizing action of saxitoxin-sensitive Na⁺ channels^{6,11}. As expected from these previous studies, saxitoxin (STX) blocked BDNF-evoked Ca²⁺ signals in cultured hippocampal neurons (*n* = 41; Fig. 1c). As it has been suggested that BDNF-evoked Ca²⁺ transients in glia are mediated by TrkB^{FL} receptors^{12,13}, we anticipated that STX would be also effective in astrocytes. STX, however, did not affect glia Ca²⁺ transients (*n* = 4; Fig. 1c). But whole-cell recordings obtained from astrocytes showed a BDNF-induced inward current (Fig. 1d; *n* = 7). The nature of

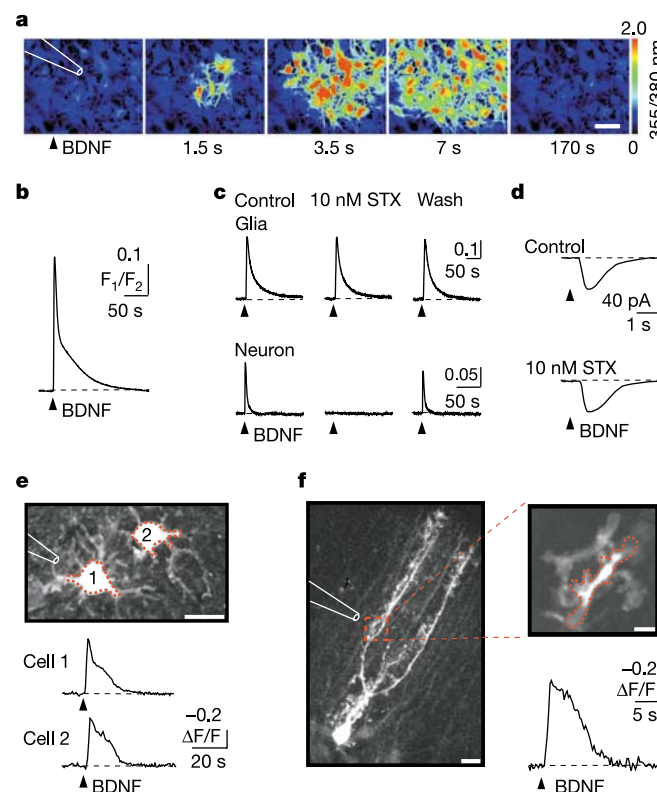


Figure 1 Ca²⁺ signalling in glia cells evoked by focal application of 0.73 nM (20 ng ml⁻¹) BDNF. **a**, BDNF-evoked Ca²⁺ wave in cultured astrocytes. **b**, Glia Ca²⁺ transient in a single astrocyte. F₁, F₂: fluorescence emission at 355 nm and 380 nm excitation, respectively. **c**, **d**, Influence of the Na⁺ channel blocker STX on Ca²⁺ transients in astrocytes and CA1 pyramidal neurons (**c**) and on a BDNF-evoked inward current in astrocytes (**d**). **e**, Image of two putative glia cells in the hippocampal stratum radiatum and BDNF-evoked Ca²⁺ signals in these cells. ΔF/F: change in fluorescence emission divided by the baseline fluorescence. **f**, Images of a Bergmann glia cell and its processes, and BDNF-evoked Ca²⁺ signal in the glia process. In **a**, **e**, **f**, the position of the BDNF ejection pipette is shown schematically; scale bars, 50 µm (**a**), 10 µm (**e**), 10 µm and 2 µm (**f**). Dotted red lines in **e** and **f** indicate the area from which the Ca²⁺ measurement was taken.

this STX-insensitive BDNF-induced current was not investigated further, but it might underlie capacitative Ca^{2+} entry¹⁴. Thus, unexpectedly, the rapid activation of glia cells by BDNF is mediated by mechanisms very different from those that activate neurons.

BDNF-evoked Ca^{2+} signals were also observed in glia cells in acute brain slices. Using two-photon Ca^{2+} imaging we identified, on the basis of morphological criteria (strong dye labelling, cell body diameter 5–10 μm , stellar morphology), putative glia cells in the stratum radiatum of hippocampal slices loaded with Fura-2-AM (Fig. 1e). Application of BDNF to single glia cells evoked Ca^{2+} signals that were similar to those in culture ($n = 25$; Fig. 1e) and spread to cells 50–100 μm away (see Supplementary Information).

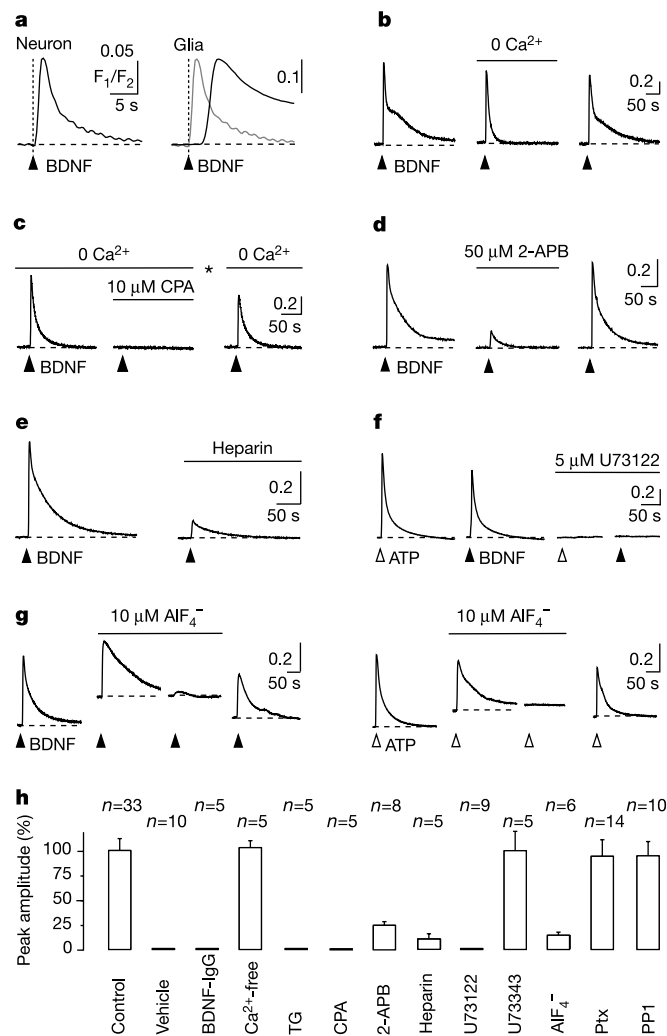


Figure 2 Mechanisms of BDNF-induced glia Ca^{2+} signals. **a**, Neuronal and glia Ca^{2+} signals have different delays. Left: neuronal Ca^{2+} transient induced by BDNF (50 ng ml⁻¹, 100 ms). Right: glia Ca^{2+} transient induced by BDNF (25 ng ml⁻¹, 100 ms). The grey trace depicts the neuronal Ca^{2+} transient shown on the left. **b**, Removal of extracellular Ca^{2+} (0 Ca^{2+}) suppresses the slow component of the recovery phase. **c–e**, The endoplasmic Ca^{2+} -ATPase inhibitor CPA (**c**), and the Ins(1,4,5) P_3 receptor antagonists 2-APB (**d**) and heparin (**e**) block Ca^{2+} signals. Asterisk in **c** indicates a 5-min period of perfusion with Ca^{2+} -containing saline to refill intracellular stores. **f**, ATP- and BDNF-evoked Ca^{2+} -signals are blocked by the PLC blocker U73122 (ATP application: 1 mM, 50 ms; BDNF: 25 ng ml⁻¹, 100 ms). **g**, Activation of heterotrimeric G proteins by AIF_4^- is sufficient to increase baseline Ca^{2+} and almost abolishes BDNF- and ATP-evoked Ca^{2+} transients. **h**, Histogram summarizing the effect of different manipulations on the peak amplitude of Ca^{2+} signals (values are the mean $\pm$ s.e.m.). TG, 10 nM thapsigargin; Ptx, 100–500 ng ml⁻¹ pertussis toxin; PP1 (src-kinase inhibitor), 20 μM .

Whole-cell recordings showed typical electrophysiological characteristics of glia cells (resting potentials around -90 mV; no spontaneous activity, inability to evoke action potentials; $n = 9$). Subsequent labelling for glia fibrillary acidic protein (GFAP) identified three out of four cells as GFAP-positive astrocytes (data not shown). The BDNF-evoked Ca^{2+} signals in glia cells *in situ* were altered neither by tetrodotoxin ($n = 4$), nor by a cocktail of glutamate receptor blockers (6-cyano-7-nitroquinoxaline-2,3-dione (CNQX), DL-2-amino-5-phosphonopentanoic acid (APV) and α -methyl-(4-carboxyphenyl)glycine (MCPG); $n = 8$; data not shown), showing that they were not due to neuronal activity or to glutamate release, respectively. BDNF-induced Ca^{2+} transients were also elicited in cerebellar Bergmann glia cells ($n = 25$; Fig. 1f).

When combining Ca^{2+} imaging with whole-cell patch-clamp recordings in cultured astrocytes, BDNF-induced Ca^{2+} signals were abolished within 15–20 min of obtaining the whole-cell mode ($n = 13$; data not shown). This run-down suggests that the BDNF-mediated response involves soluble messenger molecules that were washed out through the patch pipette. Therefore, mainly membrane-permeable probes were used with Ca^{2+} imaging to analyse the mechanisms of the BDNF-evoked glia activation.

In contrast to neurons, which respond to BDNF within less than 10 ms (ref. 5), in astrocytes the onset of the BDNF-induced Ca^{2+} signal was delayed by more than 1 s (Fig. 2a). In addition, the primary Ca^{2+} signal was independent on extracellular Ca^{2+} . Notably, however, the recovery phase was mono-exponential, and the second slow component was suppressed in the absence of extracellular Ca^{2+} (Fig. 2b). Inhibitors of endoplasmic Ca^{2+} -ATPase (cyclopiazonic acid (CPA; Fig. 2c, h) or thapsigargin (TG)) completely blocked BDNF-evoked Ca^{2+} signals (Fig. 2h). An inhibitor of inositol-1,4,5-trisphosphate (Ins(1,4,5) P_3) receptors (2-aminoethoxydiphenyl borate, 2-APB)¹⁵ reduced the Ca^{2+} transients by 75% (Fig. 2d, h). The Ins(1,4,5) P_3 receptor antagonist heparin, delivered intracellularly during a 2-min period of whole-cell patch-clamp (5 mM heparin in the pipette), also reduced the Ca^{2+} transients by about 90% (Fig. 2e, h). Performing a 2-min patch-clamp recording without heparin had no effect ($n = 5$, not shown). These data suggest that BDNF activates glia cells through the induction of Ins(1,4,5) P_3 production, followed by Ca^{2+} release from intracellular stores and Ca^{2+} entry through store-operated Ca^{2+} channels.

In astrocytes, many agonists of Ins(1,4,5) P_3 -mediated Ca^{2+} signalling, including ATP acting through G-protein-coupled P_2Y receptors, require the activation of phospholipase C (PLC)¹⁶. We therefore tested the PLC inhibitor U73122 and its inactive analogue U73343 (ref. 17). U73122 blocked both ATP- and BDNF-evoked Ca^{2+} signalling in astrocytes, whereas U73343 was ineffective (Fig. 2f, h). A general activator of heterotrimeric G proteins, AIF_4^- (ref. 18), induced an increase in baseline Ca^{2+} and prolonged Ca^{2+} transients in response to the first application of BDNF or ATP. During subsequent applications, Ca^{2+} transients were virtually abolished, indicating sustained G-protein activation by AIF_4^- and a lack of further activation by BDNF or ATP (Fig. 2g, h). Pertussis toxin, which inactivates α -subunits of the G_i - and G_o -protein families¹⁸, and PP1, a selective inhibitor of the non-receptor tyrosine kinase Src¹⁹, were ineffective (Fig. 2h). Together, these data suggest that BDNF activates PLC by activation of a G protein that is insensitive to pertussis toxin.

NT-4/5, which, like BDNF, mediates its biological actions through TrkB receptors²⁰, elicited Ca^{2+} transients that were similar to those evoked by BDNF (Fig. 3a). By contrast, NT-3, which preferentially binds to TrkC receptors²⁰, and nerve growth factor (NGF), which almost exclusively binds to TrkA receptors²⁰, did not elicit a response even at much higher concentrations ($n = 4$). This response profile indicates that the neurotrophin-induced Ca^{2+} signals in astrocytes are mediated by TrkB receptors. Unlike in neurons ($n = 9$), however, the tyrosine kinase blocker K-252a

(ref. 21), which completely blocks TrkB^{FL}-mediated excitation of neurons⁵, did not block BDNF-induced Ca²⁺ signals in glia ($n = 8$; Fig. 3b). These data indicate that the BDNF-induced Ca²⁺ transients do not require the tyrosine kinase activity of TrkB^{FL}.

We compared the expression pattern of TrkB receptors in glia cells and neurons. In cultured astrocytes at 9 days *in vitro* (DIV), an antibody specific for TrkB-T1 showed substantial staining, whereas an antibody to the C-terminal kinase domain of full-length TrkB receptors showed almost no signal (Fig. 3d). By contrast, the kinase-specific TrkB antibody, but not the TrkB-T1 antibody, strongly stained cultured hippocampal neurons (6 DIV; Fig. 3d). Note that this result does not exclude the expression of TrkB-T2 in neurons, but the lack of an antibody specific for this receptor precluded its detection at the protein level.

We used an approach based on quantitative rapid-cycle real-time polymerase chain reaction with reverse transcription (RT-PCR) to estimate²² the copy number of TrkB transcripts per amount of glia RNA. We determined very low expression of the TrkB^{FL} and the TrkB-T2 isoforms, with an estimated 1–2 copies of each per glia cell. By contrast, the expression of TrkB-T1 was more than 100-fold higher (Fig. 3c). Western blot analysis confirmed the vast abundance of TrkB-T1 protein in cultured astrocytes (data not shown). Together, these results indicate that TrkB-T1, and not TrkB^{FL}, mediates the BDNF-induced glia Ca²⁺ transients.

To verify this hypothesis, we analysed mouse mutants in which the expression of TrkB^{FL} was eliminated (trkB^{TK-/-})²³. Indeed, BDNF-induced Ca²⁺ signals were indistinguishable between astro-

cyte cultures from wild-type and those from trkB^{TK-/-} mice ($n = 17$, Fig. 4a). In addition, whereas a BDNF-induced inward current was absent in neurons derived from trkB^{TK-/-} mice, glia BDNF-induced currents were the same in wild-type and mutant mice (Fig. 4a, b). These data show that the BDNF-induced Ca²⁺ transients in glia cells do not require TrkB^{FL}.

To investigate further the involvement of TrkB-T1 receptors, we transiently transfected astrocytes with expression vectors encoding antisense TrkB-T1, TrkB-T2 or Na_v1.2 mRNA, together with a vector expressing enhanced green fluorescent protein (eGFP) as a reporter for transfection⁶. Notably, the antisense TrkB-T1 construct contained the complete sequence for TrkB-T1 including the TrkB ligand-binding domain, which is identical in all three TrkB-isoforms. In astrocytes transfected with the TrkB antisense vector, the amplitude of BDNF-evoked Ca²⁺ transients was reduced to about 45% of the amplitude of ATP-evoked Ca²⁺ signals in comparison to untransfected cultures (Fig. 4c, e). In astrocytes transfected with either the Na_v1.2 or TrkB-T2 antisense vector, the relation between the amplitude of BDNF-evoked Ca²⁺ signals and that of ATP-evoked Ca²⁺ signals was unchanged (Fig. 4e).

To verify the effectiveness of our antisense approach in reducing the protein quantities of the different TrkB isoforms, the TrkB^{FL}, TrkB-T1 or TrkB-T2 receptor was expressed in HEK-293 cells in the presence or absence of the TrkB antisense construct. Comparative immunostaining using a monoclonal antibody against the TrkB receptor domain showed that the expression of TrkB-T1 (Fig. 4d), as well as of TrkB^{FL} and TrkB-T2 (see Supplementary Information),

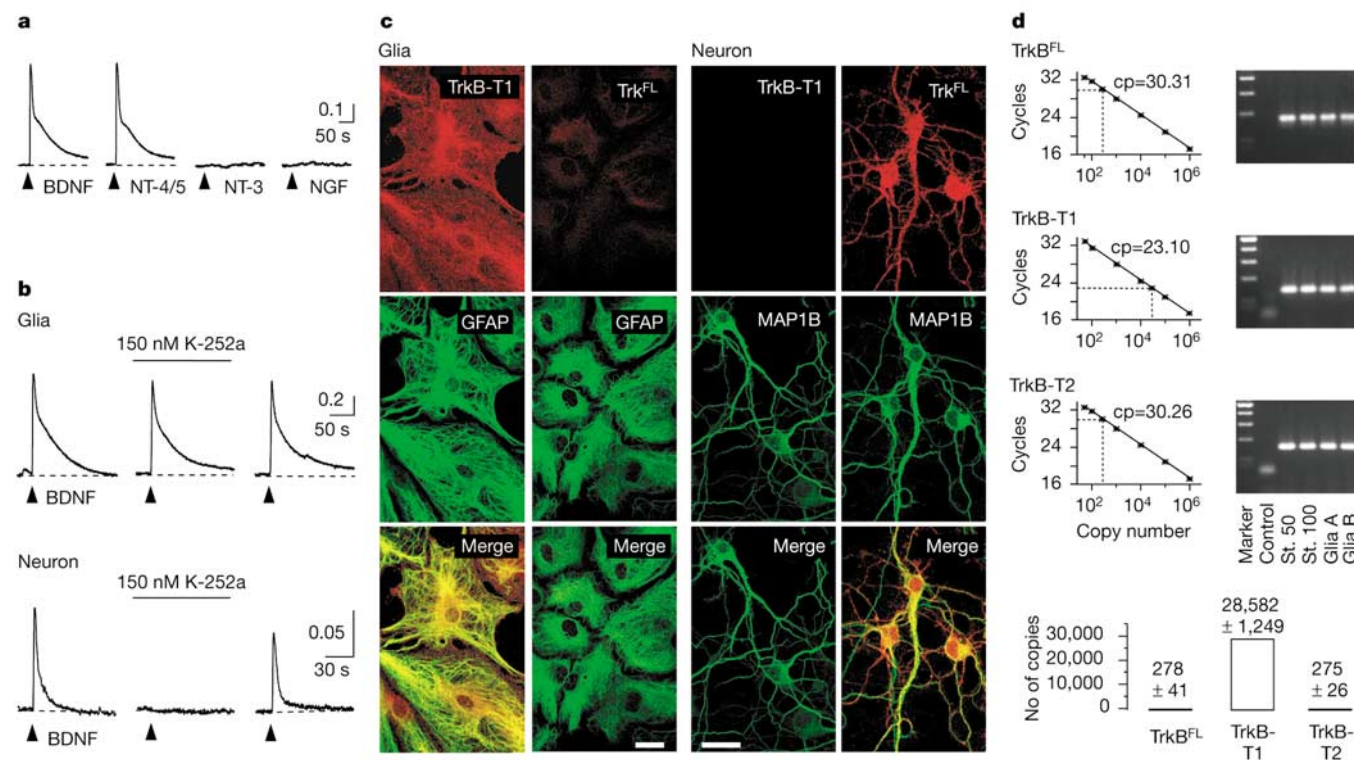


Figure 3 Distinct signalling roles of TrkB^{FL} and truncated TrkB receptors in neurons and glia. **a**, Actions of various neurotrophins in astrocytes. Application was for 150 ms at 20 ng ml⁻¹ (BDNF), 30 ng ml⁻¹ (NT-4/5), 150 ng ml⁻¹ (NT-3, NGF). **b**, Glia Ca²⁺ transients are insensitive to K-252a, whereas Ca²⁺ signals in hippocampal CA1 pyramidal neurons are blocked by K-252a. **c**, Left panels, GFAP-positive astrocytes (green) show substantial staining of TrkB-T1 (red), whereas staining of TrkB^{FL} (red) is barely above background. Right panels, MAP1B-positive cultured hippocampal neurons (green) show no staining for TrkB-T1, but stain strongly for TrkB^{FL} (red). Scale bar, 20 μ m. **d**, Quantitative RT-PCR analysis of TrkB expression in cultured astrocytes. Top left, copy

number of TrkB isoforms was determined by plotting the crossing points (cp; cycles) of the DNA standards (10⁶ to 50 copies) against their log concentration (copy number). For TrkB^{FL}, cp = 30.3 $\pm$ 0.4; y intercept = 39.2. For TrkB-T1, cp = 23.1 $\pm$ 0.2; y intercept = 39.0. For TrkB-T2, cp = 30.3 $\pm$ 0.65; y intercept = 39.1 (mean $\pm$ s.d., $n = 9$ for each isoform). Top right, agarose analysis of PCR products from the TrkB cDNA quantification. Marker, DNA markers; control, reaction without reverse transcriptase; st., standards (50 or 100 cDNA copies). Glia A and glia B are data from two different cultures. Bottom, histogram showing the mean $\pm$ s.e.m. number of transcript copies of TrkB isoforms detected from in 13 ng of glia RNA.

was almost abolished in the presence of the antisense construct.

In conclusion, our study indicates that TrkB-T1 has an independent role in glia Ca^{2+} signalling. In our model (Fig. 4f), activation of TrkB-T1 by BDNF causes activation of a G protein that stimulates PLC, production of $\text{Ins}(1,4,5)\text{P}_3$, Ca^{2+} release from stores and Ca^{2+} entry from the extracellular space. A role for truncated Trk receptors in slow cellular signalling was indicated by earlier studies, which showed that activation of TrkB-T1 and TrkB-T2 increases the release of acidic metabolites²⁴ and that truncated TrkC receptors promote neuronal differentiation²⁵. Most studies, however, have emphasized the rather passive roles of truncated TrkB receptors as 'BDNF scavengers' or inhibitors of TrkB^{FL} signalling^{2,4,20}. A direct signalling role for TrkB-T1 receptors is unexpected, because they have short intracellular carboxy-terminal tails and no intrinsic catalytic activity^{26,27}. It is therefore possible that, as for the non-catalytic receptor p75^{NTR} (ref. 28), additional interacting factors are required. So far, however, only one cytoplasmatic protein has been identified to bind TrkB-T1 (ref. 29). Alternatively, TrkB-T1

might interact with other membrane-spanning receptors. Regardless of the transduction mechanism, TrkB-T1-mediated Ca^{2+} release represents a previously unknown form of neurotrophin-evoked rapid cellular signalling and may be the dominating mechanism of neurotrophin-dependent glia activation in the brain. Because astrocytic Ca^{2+} signals modulate neuronal activity and activate other types of glia cell^{7–10}, BDNF emerges as a conceivable candidate for mediating neuron–glia interactions in the intact brain. □

Methods

Ca^{2+} imaging

Experiments were done at room temperature (22–24 °C). BDNF, NT-4/5, NT-3 and NGF (Sigma) were diluted in 'vehicle' (physiological saline containing 0.002% bovine serum albumin). Neurotrophins and ATP were applied for 20–150 ms to single cells by a Picospritzer (General Valve) coupled to standard micropipettes. $\text{trkB}^{\text{TK-/-}}$ mice²³ were purchased from Jackson Laboratories. For Ca^{2+} imaging experiments in astrocyte cultures, we list the number (*n*) of experiments, which each represent data from 20–30 individual cells. Data are expressed as the mean $\pm$ s.e.m. unless otherwise indicated. Cultured glia cells (9–15 DIV) and CA1 pyramidal neurons of hippocampal slices (P8–P10) were loaded with Fura-2-AM. Background-corrected fluorescence images were acquired by a digital imaging system (TILL Photonics) at excitation wavelengths of 355 and 380 nm. On the basis of an *in vitro* calibration, the estimated baseline Ca^{2+} concentration ranged from 50 to 100 nM and the peak amplitude of BDNF-evoked Ca^{2+} increases from 400 to 800 nM. Ca^{2+} imaging in glia cells in slices (P8–P15) was done by two-photon laser-scanning microscopy. Hippocampal glia cells were passively loaded with Fura-2-AM, and Bergmann glia cells were filled with Fura-2 via a patch pipette. Images were acquired at 4–8 Hz, corrected for background and analysed off-line with a custom-written routine in LabView (National Instruments).

Preparation of cell cultures and tissue slices and electrophysiology

Cell cultures were prepared from hippocampi of newborn Sprague-Dawley rats. Neurons were cultured on poly-L-lysine (PLL)-coated glass coverslips in Neurobasal Medium containing B27 supplement (Life Technologies). Glia cultures were seeded on poly-D-lysine (PDL)-coated glass coverslips in DMEM/F12 containing 5% fetal calf serum and G-5 supplement (Life Technologies). During Ca^{2+} -imaging experiments, cultured glia cells (confluent, 9–15 DIV) were perfused with saline containing (in mM): 125 NaCl, 3 KCl, 1.25 NaH_2PO_4 , 2 CaCl_2 , 2 MgSO_4 , 10 glucose, 10 HEPES (pH 7.4 with NaOH).

Hippocampal and cerebellar slices (300 μm) from Wistar rats ages 9–15 d and whole-cell patch-clamp recordings were done as described⁵. Combined imaging and electrophysiological studies of glia cells were done with an intracellular solution containing (in mM): 4 NaCl, 150 KCl, 10 HEPES, 0.4 GTP, 4 Mg-ATP and 0.1 mM Fura-2 (Molecular Probes); pH 7.3. For experiments in hippocampal slices with subsequent immunocytochemical staining, we used 2 mM AlexaFluor-488 (Molecular Probes) instead of Fura-2. Whole-cell recordings of neurons were done as described⁵.

Immunocytochemistry

Cells were fixed with 2% paraformaldehyde and 0.2% glutaraldehyde and incubated in blocking buffer. We used the following primary antibodies: rabbit anti-TrkB (raised against the C terminus, gp140; C14/sc-139, Santa Cruz), diluted 1:500; rabbit anti-TrkB[TK⁻] (C terminus, gp95; C13/sc-119, Santa Cruz), diluted 1:200; mouse anti-GFAP (Sigma) 1:500, and mouse anti-MAP1B (Chemicon), diluted 1:200 in blocking buffer. As secondary antibodies, Cy3-labelled goat anti-rabbit (Dianova) and Alexa488-labelled goat anti-mouse (Molecular Probes) were used. Confocal images were taken with a Fluoview 300 system (Olympus).

Quantitative rapid-cycle real-time RT-PCR

Total RNA was purified from three different glia cultures at 7–9 DIV and reverse-transcribed with Superscript (Invitrogen) and the following primers. TrkB-kinase (EMBL/GenBank m55291): 2026-for, 5'-GTTGGCGAGACATTCACAG-3'; kin-2204-rev, 5'-GGGGGTTTCAATGACAGG-3'. TrkB-T1 (EMBL/GenBank m55292): 1720-for, 5'-CGGGAGCATCTCTCGGTCT-3'; T1-1862-rev, 5'-AGGGGATCTTATGAACAAA-3'. TrkB-T2 (EMBL/GenBank m55293): 1386-for, 5'-CGGGAGCATCTCTCGGTCT-3'; T2-1538-rev, 5'-TCCACTTAAGAAGCAAAATAAGC-3'. To generate standards, cloned complementary DNAs were quantified by their absorbance at 260 nm and were then adjusted to represent from 10^6 to 50 single-stranded cDNA copies. Rapid-cycle PCR was done on a LightCycler (Roche) with an equivalent of 13 ng of glia RNA per reaction. To estimate the total amount of the different TrkB transcripts, we analysed the fluorescence of SYBR Green I, which binds to double-stranded PCR products, in the log-linear phase using the 'fit point' method with standards chosen to represent similar cDNA copy numbers²².

Antisense transfection

Rat TrkB-T1 cDNA was cloned in antisense direction in the expression vector pCAGGS (ref. 30). We cloned the corresponding cDNA of the TrkC receptor as a negative control and used a cDNA fragment of the sodium channel $\text{Na}_v1.2$ in the vector⁶ as a nonsense control. The vectors and a pCAGGS-eGFP vector were cotransfected in a molar ratio of 10:1 into glia cells at 5–6 DIV. Antisense mRNA and enhanced green fluorescent protein (eGFP) were expressed for 72–96 h before the Ca^{2+} imaging experiments were carried out.

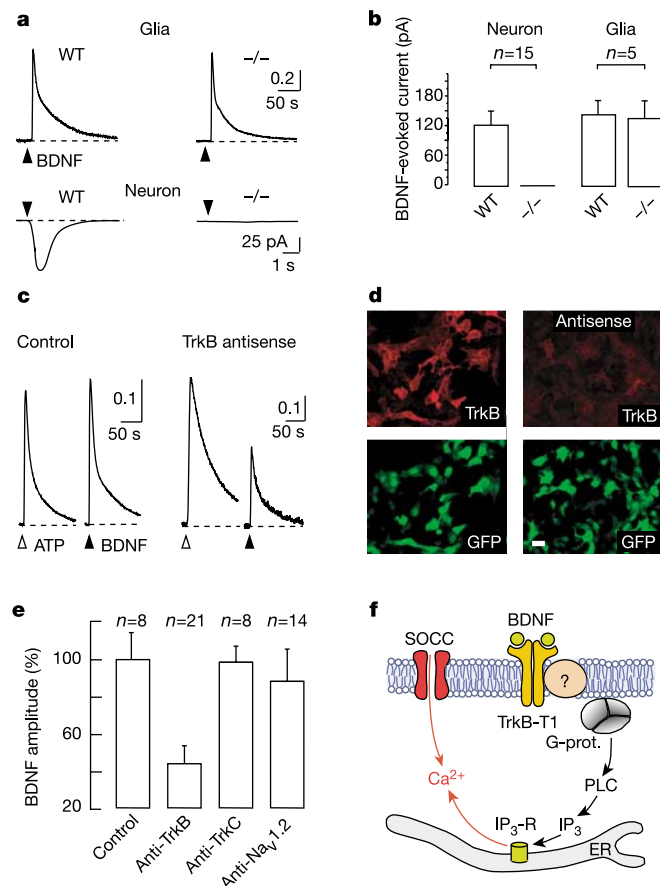


Figure 4 BDNF-evoked glia Ca^{2+} signalling through truncated TrkB-T1 receptors. **a**, Ca^{2+} transients persist in astrocytes derived from mutant mice that lack TrkB^{FL} (–/–), whereas BDNF-evoked current is absent in cultured hippocampal neurons derived from these mice. **b**, Histogram showing the mean $\pm$ s.e.m. amplitude of BDNF-evoked currents in neurons and glia cells derived from wild-type and $\text{trkB}^{\text{TK-/-}}$ (–/–) mice. **c**, Antisense TrkB expression causes a strong reduction of Ca^{2+} transients evoked by BDNF (25 ng ml^{–1}, 100 ms) as compared with Ca^{2+} transients induced by ATP (1 mM, 50 ms). **d**, Heterologous expression of TrkB-T1 receptors in HEK-293 cells is strongly reduced by co-transfection with antisense TrkB. Scale bar, 10 μm . **e**, Histogram showing mean $\pm$ s.e.m. values of the amplitude of BDNF-induced Ca^{2+} signals as compared with ATP-evoked Ca^{2+} signals in control and antisense-transfected glia cultures. **f**, Model of the mechanism of BDNF-induced Ca^{2+} signalling in glia cells. SOCC, store-operated calcium channel; G-prot., G protein.

The transfection efficiency was around 25%, as judged by the number of cells expressing eGFP. Because astrocytes show extensive intercellular Ca^{2+} signalling^{8,9}, signals from all cells in the field of view were averaged during imaging experiments.

To verify the efficiency of TrkB antisense mRNA expression, we expressed the TrkB^{FL}, TrkB-T1 or TrkB-T2 receptor in HEK-293 cells in the presence or the absence of the TrkB-T1 antisense construct. The reduction of all TrkB protein isoforms by antisense TrkB-T1 mRNA was shown by comparative immunostaining with a monoclonal antibody against the TrkB receptor domain (clone 47; BD Transduction Laboratories).

Received 24 March; accepted 6 August 2003; doi:10.1038/nature01983.

- Bonhoeffer, T. Neurotrophins and activity-dependent development of the neocortex. *Curr. Opin. Neurobiol.* **6**, 119–126 (1996).
- Thoenen, H. Neurotrophins and activity-dependent plasticity. *Prog. Brain Res.* **128**, 183–191 (2000).
- Bibel, M. & Barde, Y. A. Neurotrophins: key regulators of cell fate and cell shape in the vertebrate nervous system. *Genes Dev.* **14**, 2919–2937 (2000).
- Patapoutian, A. & Reichardt, L. F. Trk receptors: mediators of neurotrophin action. *Curr. Opin. Neurobiol.* **11**, 272–280 (2001).
- Kafitz, K. W., Rose, C. R., Thoenen, H. & Konnerth, A. Neurotrophin-evoked rapid excitation through TrkB receptors. *Nature* **401**, 918–921 (1999).
- Blum, R., Kafitz, K. W. & Konnerth, A. Neurotrophin-evoked depolarization requires the sodium channel Nav1.9. *Nature* **419**, 687–693 (2002).
- Verkhratsky, A., Orkand, R. K. & Kettenmann, H. Glial calcium: homeostasis and signaling function. *Physiol. Rev.* **78**, 99–141 (1998).
- Bezzi, P. & Volterra, A. A neuron–glia signalling network in the active brain. *Curr. Opin. Neurobiol.* **11**, 387–394 (2001).
- Haydon, P. Glia: listening and talking to the synapse. *Nature Neurosci.* **2**, 185–193 (2001).
- Newman, E. A. Calcium signaling in retinal glial cells and its effect on neuronal activity. *Prog. Brain Res.* **132**, 241–254 (2001).
- Kovalchuk, Y., Hanse, E., Kafitz, K. W. & Konnerth, A. Postsynaptic induction of BDNF-mediated long-term potentiation. *Science* **295**, 1729–1734 (2002).
- Roback, J. D., Marsh, H. N., Downen, M., Palfrey, H. C. & Wainer, B. H. BDNF-activated signal transduction in rat cortical glial cells. *Eur. J. Neurosci.* **7**, 849–862 (1995).
- Climent, E., Sancho-Tello, M., Minana, R., Barattino, D. & Guerri, C. Astrocytes in culture express the full-length TrkB receptor and respond to brain derived neurotrophic factor by changing intracellular calcium levels: effect of ethanol exposure in rats. *Neurosci. Lett.* **288**, 53–56 (2000).
- Li, H. S., Xu, X. Z. & Montell, C. Activation of a TRPC3-dependent cation current through the neurotrophin BDNF. *Neuron* **24**, 261–273 (1999).
- Maruyama, T., Kanaji, T., Nakade, S., Kanno, T. & Mikoshiba, K. 2APB, 2-aminoethoxydiphenyl borate, a membrane-penetrable modulator of Ins(1,4,5)P₃-induced Ca^{2+} -release. *J. Biochem.* **122**, 498–505 (1997).
- Idestrup, C. P. & Salter, M. W. P2Y and P2U receptors differentially release intracellular Ca^{2+} via the phospholipase C/inositol 1,4,5-triphosphate pathway in astrocytes from the dorsal spinal cord. *Neuroscience* **86**, 913–923 (1998).
- Smith, R. et al. Receptor-coupled signal transduction in human polymorphonuclear neutrophils: effects of a novel inhibitor of phospholipase C-dependent processes on cell responsiveness. *J. Pharmacol. Exp. Ther.* **253**, 688–697 (1990).
- Helm, J. B. Role of heterotrimeric GTP binding proteins in vesicular protein transport: indications for both classical and alternative G protein cycles. *FEBS Lett.* **369**, 84–88 (1995).
- Hanke, J. H. et al. Discovery of a novel, potent, and Src family-selective tyrosine kinase inhibitor. Study of Lck- and FynT-dependent T cell activation. *J. Biol. Chem.* **271**, 695–701 (1996).
- Barbacid, M. Neurotrophic factors and their receptors. *Curr. Opin. Cell Biol.* **7**, 148–155 (1995).
- Knüsel, B. & Hefti, F. K-252 compounds: modulators of neurotrophin signal transduction. *J. Neurochem.* **59**, 1987–1996 (1992).
- Rasmussen, R. in *Rapid Cycle Real-time PCR* (eds Meuer, S., Wittwer, C. & Nakagawara, K.) 21–34 (Springer, Berlin, 2001).
- Klein, R. et al. Targeted disruption of the trkB neurotrophin receptor gene results in nervous system lesions and neonatal death. *Cell* **75**, 113–122 (1993).
- Baxter, G. T. et al. Signal transduction mediated by the truncated trkB receptor isoforms, trkB.T1 and trkB.T2. *J. Neurosci.* **17**, 2683–2690 (1997).
- Hapner, S. J., Boeshore, K. L., Large, T. H. & Lefcort, F. Neural differentiation promoted by truncated trkB receptors in collaboration with p75^{NTR}. *Dev. Biol.* **201**, 90–100 (1998).
- Klein, R., Conway, D., Parada, L. F. & Barbacid, M. The trkB tyrosine protein kinase gene codes for a second neurogenic receptor that lacks the catalytic kinase domain. *Cell* **61**, 647–656 (1990).
- Middlemas, D. S., Lindberg, R. A. & Hunter, T. trkB, a neural receptor protein-tyrosine kinase: evidence for a full-length and two truncated receptors. *Mol. Cell. Biol.* **1991**, 143–153 (1991).
- Dechant, G. & Barde, Y. A. The neurotrophin receptor p75^{NTR}: novel functions and implications for diseases of the nervous system. *Nature Neurosci.* **5**, 1131–1136 (2002).
- Kryl, D. & Barker, P. A. TTIP is a novel protein that interacts with the truncated T1 TrkB neurotrophin receptor. *Biochem. Biophys. Res. Commun.* **279**, 925–930 (2000).
- Niwa, H., Yamamura, K. & Miyazaki, J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* **108**, 193–199 (1991).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank H. Thoenen for critically discussing the manuscript; T. Hunter for the TrkB cDNA isoforms; J.-i. Miyazaki for pCAGGS and pCAGGS-eGFP; M. Meyer for help with the knockout mice; and I. Schneider, R. Maul and I. Mühlhahn for technical assistance. This work was supported by grants from the Deutsche Forschungsgemeinschaft (to C.R.R. and A.K.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.R.R. (rose@lrz.uni-muenchen.de) or A.K. (konnerth@lrz.uni-muenchen.de).

Methylation at lysine 4 of histone H3 in ecdysone-dependent development of *Drosophila*

Yurii Sedkov^{1*}, Elizabeth Cho^{1*}, Svetlana Petruk¹, Lucy Cherbas², Sheryl T. Smith¹, Richard S. Jones³, Peter Cherbas², Eli Canaani⁴, James B. Jaynes¹ & Alexander Mazo¹

¹Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA

²Department of Biology, Indiana University, Bloomington, Indiana 47405-3700, USA

³Department of Biological Sciences, Southern Methodist University, Dallas, Texas 75275-0376, USA

⁴Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel

* These authors contributed equally to this work

Steroid hormones fulfil important functions in animal development. In *Drosophila*, ecdysone triggers moulting and metamorphosis through its effects on gene expression¹. Ecdysone works by binding to a nuclear receptor, EcR, which heterodimerizes with the retinoid X receptor homologue Ultraspiracle^{2,3}. Both partners are required for binding to ligand or DNA^{4–6}. Like most DNA-binding transcription factors, nuclear receptors activate or repress gene expression by recruiting co-regulators, some of which function as chromatin-modifying complexes^{7,8}. For example, p160 class coactivators associate with histone acetyltransferases and arginine histone methyltransferases⁹. The *Trithorax-related* gene of *Drosophila* encodes the SET domain protein TRR. Here we report that TRR is a histone methyltransferase capable of trimethylating lysine 4 of histone H3 (H3-K4). *trr* acts upstream of *hedgehog* (*hh*) in progression of the morphogenetic furrow, and is required for retinal differentiation. Mutations in *trr* interact in eye development with *EcR*, and *EcR* and TRR can be co-immunoprecipitated on ecdysone treatment. TRR, *EcR* and trimethylated H3-K4 are detected at the ecdysone-inducible promoters of *hh* and *BR-C* in cultured cells, and H3-K4 trimethylation at these promoters is decreased in embryos lacking a functional copy of *trr*. We propose that TRR functions as a coactivator of *EcR* by altering the chromatin structure at ecdysone-responsive promoters.

TRR¹⁰ was identified as a protein that is closely related to the well-known *HOX* gene activator Trithorax (TRX). An antibody to TRR recognized a protein band of the expected size (relative molecular mass (*M_r*) 260) in nuclear extracts from embryos (Fig. 1b), suggesting that TRR is a nuclear protein. Nuclear localization of TRR was confirmed by immunohistochemical staining of salivary glands from third instar larvae (Fig. 1c). TRR contains several protein domains that are conserved among transcriptional regulators (Fig. 1a), including a SET domain. This ancient domain is known to bind to several proteins, including histones¹¹, and to have histone methyltransferase (HMTase) activity (reviewed in ref. 12). Indeed, TRR from nuclear extracts bound to a column containing the four core histones; and the isolated SET domain of TRR, like the closely related TRX SET¹¹, bound to histones H3 and H4 (Supplementary Fig. 1a, b).

Purified TRR SET showed histone-H3-specific HMTase activity *in vitro* (Fig. 1d, e). It also methylated a histone H3 amino-terminal peptide that was either unmodified or trimethylated at lysine 9 (K9), but it did not methylate a similar peptide that was already trimethylated at K4 (Fig. 1d). To test whether other lysine residues known to be methylated in histone H3 are methylated by TRR SET, we tested histones mutated at K4, K27, K36 and K79 (ref. 13). Only mutation

of K4 affected the ability of histone H3 to be methylated (Fig. 1e). K4 specificity was confirmed by a failure of antibodies specific for H3-MeK27 and H3-MeK36 to recognize the product of TRR methylation (data not shown). K4 was directly confirmed as a target of TRR methylation in histone H3 by Edman degradation of H3 methylated *in vitro* by TRR SET (Fig. 1f). Further analysis showed that TRR SET can trimethylate at H3-K4 (Fig. 1g). Because trimethylated H3-K4 is associated with high transcriptional activity¹⁴, these results indicate that TRR may function as a transcriptional coactivator.

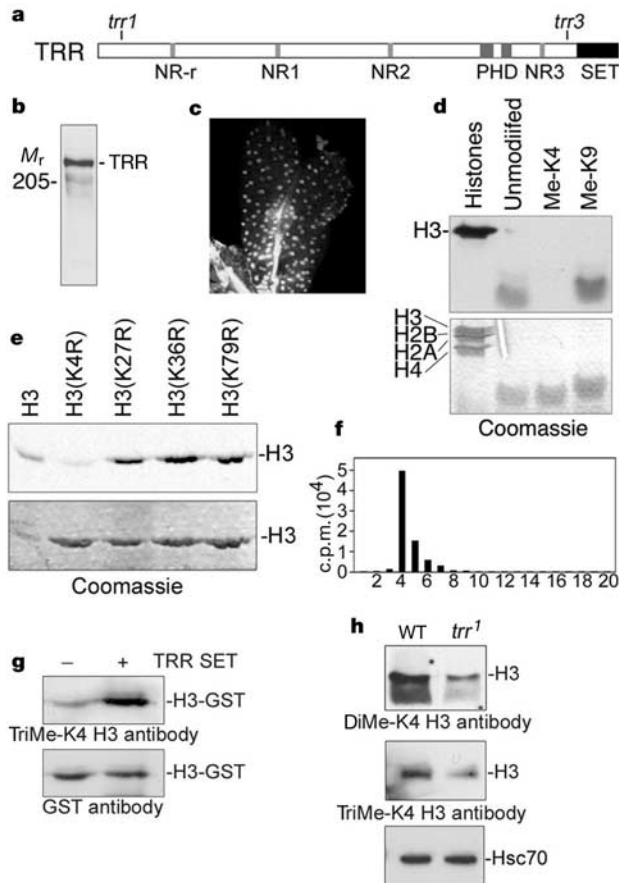


Figure 1 TRR is a nuclear protein that methylates H3-K4. **a**, Structure of TRR. Shown are NR1–NR3 and NR-R, the LXXLL and the LLXXL nuclear receptor interacting motifs, respectively. PHD, PHD finger domains. Point mutations in *trr*¹ and *trr*³ are indicated; both mutations lead to a stop codon. **b**, Detection of TRR in embryo extracts by western blotting. **c**, Detection of TRR in nuclei of third instar larval salivary glands. **d–f**, Specificity of the TRR SET domain for H3-K4 methylation. **d**, Of the four core histones, TRR SET specifically methylates histone H3. ^3H -methylated histones were visualized by SDS–PAGE and autoradiography. TRR SET methylated the unmodified N-terminal peptide of histone H3 (unmodified) and an N-terminal peptide trimethylated at K9 (Me-K9), but not a peptide trimethylated at K4 (Me-K4). **e**, Equal amounts of wild-type and several lysine-mutated derivatives of histone H3 were tested (as described in **d**) for their ability to be methylated by TRR SET. Mutated lysines are indicated. Input controls stained by Coomassie blue are shown in the bottom panels of **d** and **e**. **f**, The ^3H -methylated wild-type histone H3 from **d** was subjected to Edman degradation, and ^3H released at each cycle was measured by scintillation counting. **g**, Trimethylation of histone H3-K4 by TRR SET. After methylation of recombinant GST–histone H3 by purified TRR SET, the sample was separated by SDS–PAGE and probed with either antibodies against histone H3 trimethylated at K4 (top) or antibodies against GST (bottom). **h**, Western blot analysis of the methylation state of H3-K4 in extracts from wild-type and *trr*¹ mutant embryos. In each lane, 15 μg of total protein extract was loaded; the bottom panel shows the loading control using antibody against Hsc70.

Ash1 and TRX are also H3-K4 HMTases in *Drosophila*^{15,16}. TRR seemed to be more abundant in embryos than either Ash1 or TRX (data not shown), and it was associated with many more sites on salivary gland polytene chromosomes (Fig. 2c). To test the magnitude of TRR's contribution to H3-K4 methylation, we compared the amounts of H3-MeK4 in extracts from wild-type embryos with those from embryos homozygous mutant for the *trr*¹ null allele¹⁰. Mutation of *trr* markedly reduced the quantity of both di- and trimethylated H3-K4 (Fig. 1h). By contrast, we detected only a small decrease in methylated H3-K4 in embryos homozygous for a *trx* null mutation (*trx*^{B11}, data not shown). We conclude that TRR is a major H3-K4 HMTase in *Drosophila*.

Unlike TRX, TRR contains four sequences that match the nuclear-receptor-interacting motifs¹⁷ LXXLL or LLXXL ('NR' in Fig. 1a). We tested each TRR NR region for its ability to interact with the ecdysone receptor *in vitro*. The NR2 domain of TRR, which contains an LXXLL motif, bound to full-length EcR and Ultra-spiracle (USP) proteins; although EcR and USP are considered to be obligate heterodimer partners, either protein bound to the TRR NR2 domain alone (Fig. 2a). The three other NR domains of TRR showed very little if any interaction with EcR and USP (data not shown). An association between TRR and the ecdysone receptor *in vivo* was shown by co-immunoprecipitation from embryo extracts (Fig. 2b), in which TRR was associated with both components of the ecdysone receptor.

We confirmed the association of TRR and EcR *in vivo* by two additional experiments. Like TRX¹⁸, TRR bound to several sites on salivary gland polytene chromosomes; immunostaining of polytene chromosomes showed an almost complete colocalization of TRR with EcR-B1 (Fig. 2c), the predominant EcR isoform of the salivary gland. Although the stained bands were numerous, they constituted only a few of the bands visible by DNA staining (blue). By contrast, there was little overlap between the staining for TRR and either TRX or Polycomb (data not shown). In the eye imaginal disc, TRR was expressed in a pattern essentially identical to those of USP and EcR-A, the predominant EcR isoform of eye discs (Fig. 2d). As reported for USP¹⁹, both TRR and EcR-A were expressed primarily in the region posterior to the morphogenetic furrow and in smaller amounts in a row of cells just anterior to the morphogenetic furrow, and all three proteins were absent from most of the region anterior to the morphogenetic furrow. Thus, in both the salivary gland and the eye disc, there is marked spatial colocalization of TRR and EcR–USP.

Having established a physical association between TRR and the ecdysone receptor, we tested for a functional association *in vivo*, concentrating on the developing compound eye. Ecdysone and its receptor have complex roles in both progression of the morphogenetic furrow^{19,20} and post-furrow differentiation²¹. We showed that TRR is required for eye development by examining the phenotype of somatic *trr*¹ clones in the adult eye. Two abnormalities were visible in the clonal patches: a notching of the anterior boundary of the eye, interpretable as a retardation of morphogenetic furrow progression; and a 'rough eye' phenotype, attributable to defects in photoreceptor differentiation (Fig. 2e).

Three approaches showed that *trr* interacts genetically with genes encoding components of the ecdysone receptor and the ecdysone synthesis pathway. First, heterozygosity for the null allele *trr*¹, similar to the temperature-sensitive *ecdysoneless* allele *ecd*¹ (at its partially restrictive temperature, 25°C), enhanced the furrow-stop phenotype of the *B*¹ mutation (Supplementary Fig. 2a, b). Introducing an extra copy of *trr*¹ as a transgene¹⁰ reversed the *trr*¹ effect (Supplementary Fig. 2a). Flies with a combination of *ecd*¹, *trr*¹ and *B*¹ showed an additional defect in morphogenetic furrow progression, particularly in the ventral part of the eye (Supplementary Fig. 2b, bracket). *trr*³, in which the SET domain is deleted, was similar to *trr*¹ in its interaction with *B*¹ (data not shown). These genetic interactions are consistent with a role for TRR as a coactivator for the ecdysone receptor.

Second, expression of the dominant-negative mutant EcR-F645A in the post-morphogenetic furrow retinal epithelium by the *GMR-GAL4* driver led to severe defects in eye differentiation (Fig. 2f and Supplementary Fig. 2d) and in a marked lethality before adult eclosion (ref. 10 and Table 1). Under the conditions used, lethality was apparently due largely to desiccation, presumably caused by defective cuticle secretion. The lethality associated with EcR-F645A was strongly enhanced in *trr*¹ and *trr*³ heterozygotes (Table 1). Inclusion of an extra copy of *trr*⁺ partially suppressed both the lethality and the eye phenotype (Table 1 and Fig. 2f). A mutation in *Smrter* (*Smr*), which encodes a known co-repressor for the ecdysone receptor²², showed an interaction with EcR-F645A similar to that observed with an extra copy of *trr*⁺ (Table 1 and

Fig. 2f). Thus, *trr* and *Smr* function antagonistically, which is again consistent with a role for TRR as a coactivator for the ecdysone receptor.

Third, heterozygosity for the EcR null allele *EcR*^{M554fs} noticeably suppressed the furrow-stop phenotype of *trr*¹/*B*¹ flies (Fig. 2g, arrows). Homozygotes for the viable allele *trr*⁴, like the *trr*¹ null clones, showed both defects in morphogenetic furrow progression and disorganization of the ommatidial array (Fig. 2h; *trr*⁴ did not complement *trr*¹ (data not shown) and is associated with insertion of a P-element in the *trr* promoter region). Both *trr*⁴ phenotypes were suppressed by heterozygosity for *EcR*^{k06210}, an isoform-A-specific hypomorphic mutation, which by itself caused no visible eye phenotype (Fig. 2h, left).

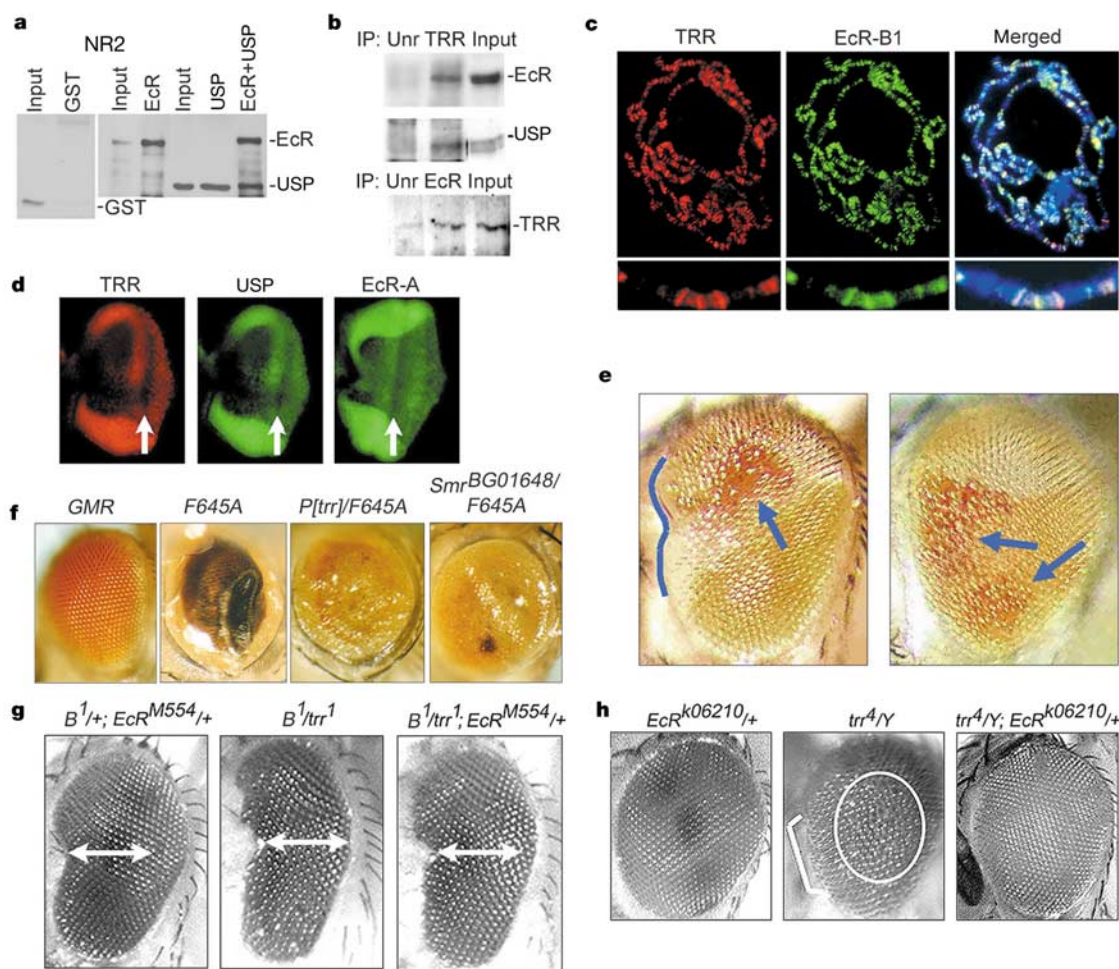


Figure 2 TRR interacts genetically and associates *in vitro* and *in vivo* with components of the ecdysone receptor complex. **a**, TRR NR2 binds EcR and USP. Bacterial extracts containing GST, GST–EcR or GST–USP were passed through a column containing the HA-tagged NR2 domain of TRR (Fig. 1a). Bound material was eluted and resolved by SDS–PAGE. **b**, Association of TRR with EcR and USP *in vivo*. Embryonic nuclear extracts were subjected to immunoprecipitation (IP) with an unrelated anti-GST, anti-HA or anti- β -Gal antibody, or with anti-TRR and anti-EcR antibodies. Immunoprecipitates were analysed by western blotting with the antibodies indicated on the right. **c**, Top, immunostaining of salivary gland polytene chromosomes with antibodies against the EcR-B1 isoform (green) and TRR (red), followed by Hoechst staining for DNA (blue). Bottom, high-magnification images of part of a chromosome. **d**, Left and middle, double immunostaining of an eye imaginal disc for TRR (red) and USP (green); right, immunostaining of a similar disc for EcR-A. Arrows indicate the morphogenetic furrow. Note that most of the region anterior to the morphogenetic furrow (left) is not stained with

any of the antibodies. **e**, Clonal analysis of *trr*. Defects in eye structure in *trr*¹/*trr*¹ clones are marked by *w*⁺ (red); blue line on the left indicates a defect due to incomplete morphogenetic furrow progression; blue arrows indicate defects in ommatidial organization. **f**, Interaction of *trr* and *Smr* with a dominant-negative EcR mutant in the post-furrow retinal epithelium. Representative eyes were photographed through a dissecting microscope. *GMR*, *GMR* driver alone. *F645A*, *GMR* driver and EcR-F645A responder; *P[trr]/F645A*, *GMR* driver and EcR-F645A responder plus extra copy of *trr*; *Smr*^{BG01648}/*F645A*, *GMR* driver and EcR-F645A responder, hemizygous for a weak mutation in *Smr* (see also Supplementary Fig. 2b). **g**, Suppression of the *B*¹/*trr*¹ phenotype by *EcR*^{M554}. Double-headed arrow (same length in each panel) illustrates differing extents of the defect in morphogenetic furrow progression in different mutant combinations. **h**, Complete rescue of the *trr*⁴ mutant phenotype by *EcR*^{k06210}. In *trr*⁴/*Y*, the bracket indicates defects caused by incomplete morphogenetic furrow progression; the oval encloses defects in ommatidial structure.

In contrast to the results from the first two approaches, these interactions between *trr* and a reduction in EcR function indicate that TRR may function in an opposite direction to EcR. These apparently paradoxical results are readily explained by reference to the fact that ecdysone converts its receptor from a transcriptional repressor to an activator. Both *ecd¹* and EcR-F645A specifically abolish the activation function without altering repression²³, and both enhance the phenotype of *trr* mutations, consistent with a coactivator function for TRR. Mutations in EcR affect both receptor functions and suppress the phenotype of *trr* mutations. This suggests that EcR-mediated repression has an essential, dose-sensitive role that can be partially overcome by a decrease in coactivator (TRR) titre. Similarly complex effects in eye development have been seen for mutations in *usp*²⁴.

We sought to clarify the role of TRR in furrow progression. Our analyses of genetic interactions between *trr* and both *hedgehog* (*hh*) and its downstream target *decapentaplegic* (*dpp*) indicated that TRR has a role in Hedgehog (Hh) signalling (Fig. 3). Like *B¹*, *trr¹* was found to be a dominant enhancer of the eye-specific mutation *dpp^{blk}* (ref. 25 and Fig. 3a). Hh is involved in morphogenetic furrow progression and in post-furrow photoreceptor differentiation²⁶, and *trr* mutations altered *hh* phenotypes in both processes. Although a reduction in *trr* dose had no apparent effect on either aspect of the *hh¹* phenotype (data not shown), an extra copy of *trr* visibly suppressed the furrow-stop phenotype and restored normal ommatidial differentiation in *hh¹* homozygotes. Flies heterozygous for both *trr¹* and *hh^{13C}* also showed a rough-eye phenotype that was not produced by either mutation alone (Fig. 3b).

Because heterozygosity for *trr¹* enhanced the morphogenetic furrow progression defect of *B¹* flies (Supplementary Fig. 2a), we tested whether it also enhanced the reduction in *dpp* and *hh* expression caused by *B¹*. Heterozygosity for *trr¹* alone had no effect on eye morphology or on *hh* and *dpp* expression (data not shown). *B¹* caused a decrease in expression of *dpp* in the middle of the eye disc, whereas *dpp* expression was completely abolished in this region in *B¹/trr¹* transheterozygotes (Fig. 3c). *trr¹* also clearly enhanced the deficit in *hh* expression caused by *B¹* (Fig. 3d). The requirement of *trr* for *hh* expression was further shown by the considerable decrease in *hh* expression in the *trr* null clones posterior to the morphogenetic furrow (Fig. 3e). The spatial pattern of *trr* expression is consistent with TRR having a direct role in *hh* expression (Figs 2d and 3f), whereas there is little if any overlap between the areas of expression of *trr* and *dpp* (Fig. 3f). Because *hh* functions upstream of *dpp* in directing morphogenetic furrow progression, these results indicate that TRR may positively regulate *hh* transcription, and that a decrease in *hh* expression in *trr* mutants may lead to a decrease in the expression of *dpp*.

Alternatively, TRR might upregulate *dpp* directly, in addition to its indirect effect through *hh*. These possibilities can be distinguished

by use of the dominant hypermorphic mutation *ro^{Dom}*: Hh stimulates *ro* expression, and *ro^{Dom}* interferes with *hh* induction of *dpp*^{27–29}. If *trr* directly upregulates *dpp*, then an extra copy of *trr* would be expected to suppress the *ro^{Dom}* phenotype. If the effect of TRR on *dpp* expression is mediated completely by *hh*, then an extra copy of *trr* would be expected to enhance the *ro^{Dom}* phenotype. We found that the latter is the case (Supplementary Fig. 2b). Therefore, the effects of TRR on *dpp* expression are likely to be predominantly a secondary consequence of its effects on *hh* expression.

The above experiments lead to the hypothesis that TRR functions as a coactivator for the ecdysone receptor, mediating ecdysone activation of transcription by modifying chromatin structure at target promoters. The results also indicate that *hh* may be a direct TRR target during eye development. This hypothesis predicts an ecdysone-dependent association of TRR with the ecdysone receptor and with ecdysone-responsive promoters, and an ecdysone induction of H3-K4 trimethylation at ecdysone-induced promoters. We tested these predictions as follows.

Fractionation of an embryo extract showed that TRR was present in very large complexes, $M_r \approx 2,000,000$ (2,000K), whereas EcR migrated as a broad band, predominantly at a much lower M_r . Treatment of the extract with ecdysone caused the TRR peak to become confined to a higher M_r range and caused a marked shift in migration of EcR to higher M_r , so that roughly half of the EcR co-migrated with TRR (Fig. 4a). Similar results were obtained with S2 cell extracts (data not shown). Antibody specific for EcR could precipitate TRR more efficiently from the 2,000K fraction of the hormone-treated embryonic extract than from the untreated extract (Fig. 4b). Thus, TRR and EcR are components of a hormone-dependent complex *in vivo*. Although the presence of USP in this complex has not been shown, the hormone dependence of the EcR–TRR interaction indicates that TRR may have a direct role in mediating transcriptional induction by the ecdysone receptor.

We then used S2 cells to examine the chromatin structure of two putative target promoters. We examined *hh*, which was suggested by our results to be a target of TRR activation in the eye disc, and *BR-C*, a gene whose transcription is ecdysone-responsive in many tissues, including the eye^{20,24}. Transcription of both genes was induced by ecdysone in S2 cells (Fig. 4c). Chromatin immunoprecipitation (ChIP) assays showed that ecdysone treatment caused a marked increase in the association of both promoters with EcR, TRR and H3 trimethylated at K4 (Fig. 4d). A causal connection between the presence of TRR and histone methylation at these promoters was indicated by similar assays comparing extracts from wild-type embryos and from embryos homozygous for strong mutations at *trr*: transcription and the amount of H3-K4 trimethylation at both promoters was significantly reduced in embryos homozygous for either *trr¹* or *trr³* (Fig. 4e, f). In addition, because the SET domain

Table 1 Lethality in flies expressing EcR-F645A in the eye is enhanced by mutation in *trr* and suppressed by an extra copy of *trr* or mutation in *Smr*

Genotype	Classes of progeny that were scored	Percentage (number) of mutant chromosomes recovered	Percentage (number) of balancer chromosomes recovered
<i>trr¹</i>	Females: <i>trr</i> [OR] FM7C/+; UAS-EcR-F645A/GMR-GAL4	3% (4)	97% (120)
<i>trr³</i>	Females: <i>trr</i> [OR] FM7C/+; UAS-EcR-F645A/GMR-GAL4	<2% (0)	>98% (41)
<i>P[trr]</i>	Females: UAS-EcR-F645A/GMR-GAL4; <i>P[trr]</i> [OR] TM3/+	63% (19)	37% (11)
	Males: UAS-EcR-F645A/GMR-GAL4; <i>P[trr]</i> [OR] TM3/+	82% (58)	18% (13)
<i>Smr^{BG01648}</i>	Females: <i>Smr</i> [OR] FM7C/+; UAS-EcR-F645A/GMR-GAL4	59% (377)	41% (266)
	Males: <i>Smr</i> [OR] FM7C/Y; UAS-EcR-F645A/GMR-GAL4	94% (530)	6% (33)

Data were pooled from at least ten replicates for each of the following crosses. *trr¹*: female *trr¹/FM7C*; UAS-EcR-F645A/CyO × male GMR-GAL4. As *trr¹* is homozygous lethal (homozygous lethal in males), only female progeny were scored. *Smr^{BG01648}*: female *Smr^{BG01648}/FM7C*; UAS-EcR-F645A × male GMR-GAL4. *Smr^{BG01648}* is a homozygous-viable presumptive hypomorph. *Smr^{BG01648}/Y* males are therefore predicted to have a much lower titre of SMR than are *Smr^{BG01648}/+* females. *P[trr]*: UAS-F645A/SM5; *P[trr]/TM3* × GMR-GAL4. Males produce a higher titre of TRR than do females because of dosage compensation effects on the X-derived *trr* transgene. Under the conditions of our assay, expression of EcR-F645A in the GMR domain causes roughly 95% lethality before adult eclosion. Data in the balancer column are consistent with this value in all cases except for female progeny of the *Smr* cross, in which the viability of balancer-containing flies was strongly increased by an unknown mechanism. In all cases, viability was near 100% if either UAS-EcR-F645A or GMR-GAL4 was absent (data not shown).

of TRR is deleted in *trr*³ (Fig. 1a), this indicates that TRR SET may be responsible for the ecdysone-induced H3-K4 methylation at these promoters.

Our experiments show that TRR is a major *Drosophila* HMTase that binds to EcR in an ecdysone-dependent manner and causes H3-K4 trimethylation at ecdysone-inducible promoters. In the developing compound eye, TRR has a crucial role in progression

of the morphogenetic furrow and in post-furrow differentiation of the retinal epithelium, and one of its direct targets is the *hh* promoter. Taken together, our results imply that TRR is a coactivator for the ecdysone receptor. □

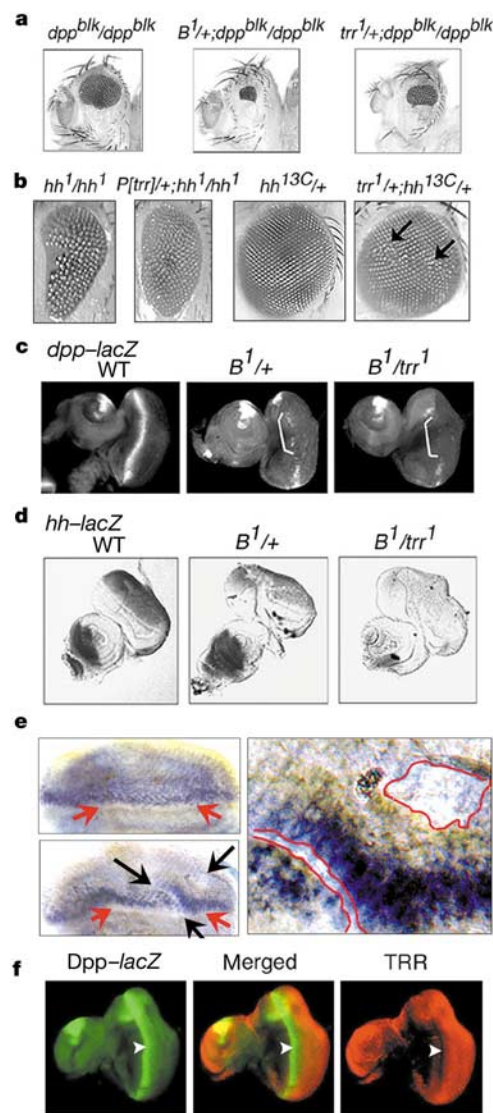


Figure 3 The role of *trr* during eye development. **a**, Enhancement of the *dpp*^{blk} phenotype by *B*¹ and *trr*¹. **b**, Suppression of the *hh*¹/*hh*¹ 'furrow-stop' phenotype by the *trr*⁺ transgene (left), and mutant phenotype caused by the transheterozygous combination of *hh*^{13C} and *trr*¹ (right, arrows indicate defects seen only in the double heterozygote). **c**, **d**, Effects of *B*¹ and transheterozygous *B*¹ with *trr*¹ (*B*¹/*trr*¹) on the expression in eye discs of *dpp-lacZ* (from transposon BS3.0 *dpp-lacZ*) and *hh-lacZ* (from transposon *hh*^{P30}), assayed by immunostaining with LacZ antibodies. In **c**, brackets indicate the region that is differentially affected by *trr*¹ addition. **e**, Expression of *hh* RNA (blue) is decreased in *trr* null clones, indicated by the absence of the green fluorescent protein marker (light brown). Two eye imaginal discs from the same larva are shown; in the bottom disc, *trr* null clones posterior to the morphogenetic furrow (red arrows) are indicated by black arrows. A higher magnification image of the *trr* null clones is shown on the right. **f**, Expression of *dpp-lacZ* and TRR in wild-type eye discs, assayed by double immunostaining. Note that TRR is not expressed in the morphogenetic furrow (indicated by arrows).

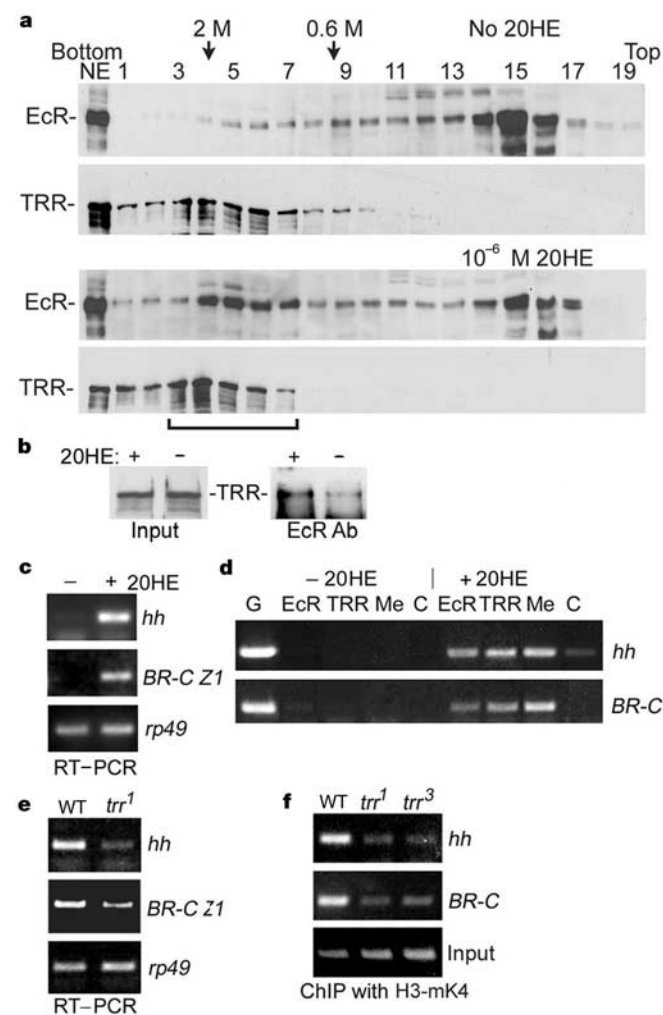


Figure 4 Molecular effects of ecdysone. **a**, Glycerol gradient analysis of TRR- and EcR-containing complexes in embryo extracts. An extract from wild-type embryos was divided into two aliquots, one of which (bottom two panels) was treated for 1 h with 10⁻⁶ M 20HE. Each aliquot was separated by centrifugation through a glycerol gradient. Fractions were analysed by western blotting. Approximate sizes, based on calibration using patterns of known complex components, are indicated at the top. **b**, Co-immunoprecipitation of glycerol gradient fractions. Right, the bracketed fractions in **a**, from either untreated (–) or 20HE-treated (+) extract, were precipitated with antibodies against EcR-C and analysed with antibodies against TRR. Left, input amounts of TRR. **c**, RT-PCR analysis of *hh* and *BR-C* RNA in uninduced (–) and 20HE-treated (+) S2 cells. The same sets of primers, specific to the 3' regions of the *hh* and *BR-C* Z1 transcription units, were used in **c** and **e**. Primers specific to the constitutively expressed *rp49* were tested on the same RNA samples as a control. **d**, ChIP analysis of the *hh* and *BR-C* Z1 promoters in S2 cells using antibodies against EcR-B1, TRR and H3 trimethylated on K4 (Me). +20HE, addition of 20HE; G, total genomic DNA; C, control ChIP without antibody. Primers used were specific to the promoter regions of *hh* and *BR-C* Z1. **e**, RT-PCR analysis of *hh*, *BR-C* Z1 and *rp49* (as a control) in RNA isolated from wild-type and *trr*¹ mutant embryos, as indicated. **f**, ChIP analysis of the methylation state of K4 in histone H3 in wild-type and *trr* mutant embryos, using antibodies specific to H3 trimethylated at K4. 'Input', control PCR amplification with the *hh* primer set using 10% of the starting material before immunoprecipitation.

Methods

Fly strains, genetic and clonal analyses

Details of the genetic manipulations are given in the Supplementary Information.

Histology and immunocytochemistry

The anti-TRR polyclonal antiserum T1 was raised in rabbits (Cocalico Biological) against a bacterially expressed 6×His fusion protein containing TRR residues 568–828. The T1 antiserum was affinity purified, and its specificity was confirmed by the absence of staining in homozygous *trr*¹ null mutant embryos (data not shown). Monoclonal antibodies to EcR-A (15G1a) and EcR-B1 (AD4.4) were a gift from C. Thummel. Immunostaining of polytene chromosomes and eye discs is described in the Supplementary Information.

Protein analysis

Western blotting, co-immunoprecipitation, *in vitro* binding assays and size fractionation of nuclear extract were done by standard procedures (see Supplementary Information) with the following reagents: antibodies against histone H3 either dimethylated at K4 (Upstate Biotechnology) or trimethylated at K4 (Abcam); T1 antibodies against TRR; antibodies against glutathione *S*-transferase (GST; diluted 1:1,000; Amersham Pharmacia Biotech); DDA2.7 monoclonal antibodies against the EcR common region (1:1,000; a gift from C. Thummel); and monoclonal antibodies against USP (1:50; a gift from F. Kafatos).

Semi-quantitative RT-PCR

Schneider S2 cells were grown in the presence of 10^{−6} M 20-hydroxy-ecdysone (20HE) for 24 h. RNA was extracted from roughly 10⁶ S2 cells or 50 embryos with a High Pure RNA Isolation kit (Roche), and reverse transcription was done with random hexamers and AMV reverse transcriptase. To normalize the relative amounts of *hh* and *BR-C Z1* RNAs, we first analysed samples by PCR with *rp49*-specific primers. Semi-quantitative PCR was achieved by preparing serial dilutions of input DNA. All samples were assayed in the range where the PCR products showed a linear correlation with input DNA. We used the following primers. For *rp49*: sense, 5′-cgg atc gat atg cta agc tg-3′; antisense, 5′-gaa cgc agg cga cct tgg ggg-3′. For *hh*: sense, 5′-tga tcc caa cga tcc tgg-3′; antisense, 5′-gta aca cgc tct gtt ttc t-3′. For *BR-C Z1*: sense, 5′-ccg acc aat tgc aaa gat ct-3′; antisense, 5′-gta cat gga tct cat cgt ca-3′.

ChIP assays

Experiments were done by standard protocols and a ChIP Assay kit (Upstate Biotechnology), using the same starting material as in the RT-PCR experiments. For normalization, the unprecipitated material was analysed by PCR using the primer set specific to the *hh* promoter. We used the following primers. For the *hh* promoter (−564 to −173): sense, 5′-tgg atg aaa gtt gtt gcc t-3′; antisense, 5′-acc att acc tga ggg gca tat-3′. For the *BR-C Z1* promoter (−554 to −120): sense, 5′-tcc att tgc gat gct gtt gct-3′; antisense, 5′-agg ttc act gca gtt cag agt-3′.

Histone methyltransferase assay

The SET domain of TRR (residues 2,199–2,410) was expressed in *Escherichia coli* as a GST fusion, and assayed for HMTase activity as described¹². Recombinant mutant histones H3 were a gift from Y. Zhang. Edman degradation was done at the KCC Protein Facility on a 477A sequencer (Applied Biosystems).

Received 23 May; accepted 15 September 2003; doi:10.1038/nature02080.

- Riddiford, L. M. Hormone receptors and the regulation of insect metamorphosis. *Receptor* **3**, 203–209 (1993).
- Oro, A. E., McKeown, M. & Evans, R. M. Relationship between the product of the *Drosophila ultraspiracle* locus and the vertebrate retinoid X receptor. *Nature* **347**, 298–301 (1990).
- Koelle, M. R. *et al.* The *Drosophila EcR* gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily. *Cell* **67**, 59–77 (1991).
- Thomas, H. E., Stunnenberg, H. G. & Stewart, A. F. Heterodimerization of the *Drosophila* ecdysone receptor with retinoid X receptor and ultraspiracle. *Nature* **362**, 471–475 (1993).
- Yao, T. P., Segaves, W. A., Oro, A. E., McKeown, M. & Evans, R. M. *Drosophila ultraspiracle* modulates ecdysone receptor function via heterodimer formation. *Cell* **71**, 63–72 (1992).
- Yao, T. P. *et al.* Functional ecdysone receptor is the product of EcR and Ultraspiracle genes. *Nature* **366**, 476–479 (1993).
- Hermanson, O., Glass, C. K. & Rosenfeld, M. G. Nuclear receptor coregulators: multiple modes of modification. *Trends Endocrinol. Metab.* **13**, 55–60 (2002).
- Stallcup, M. R. Role of protein methylation in chromatin remodeling and transcriptional regulation. *Oncogene* **20**, 3014–3020 (2001).
- Lee, Y. H., Koh, S. S., Zhang, X., Cheng, X. & Stallcup, M. R. Synergy among nuclear receptor coactivators: selective requirement for protein methyltransferase and acetyltransferase activities. *Mol. Cell. Biol.* **22**, 3621–3632 (2002).
- Sedkov, Y. *et al.* Molecular genetic analysis of the *Drosophila trithorax-related* gene which encodes a novel SET domain protein. *Mech. Dev.* **82**, 171–179 (1999).
- Katsani, K. R., Arredondo, J. J., Kal, A. J. & Verrijzer, C. P. A homeotic mutation in the *trithorax* SET domain impedes histone binding. *Genes Dev.* **15**, 2197–2202 (2001).
- Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
- Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).
- Santos-Rosa, H. *et al.* Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407–411 (2002).
- Beisel, C., Imhof, A., Greene, J., Kremmer, E. & Sauer, F. Histone methylation by the *Drosophila* epigenetic transcriptional regulator Ash1. *Nature* **419**, 857–862 (2002).
- Smith, S. T. *et al.* Modulation of heat shock gene expression by the TAC1 chromatin modifying complex. *Nature Cell Biol.* (submitted).

- Heery, D. M., Kalkhoven, E., Hoare, S. & Parker, M. G. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* **387**, 733–736 (1997).
- Kuzin, B., Tillib, S., Sedkov, Y., Mizrokhi, L. & Mazo, A. The *Drosophila trithorax* gene encodes a chromosomal protein and directly regulates the region-specific homeotic gene fork head. *Genes Dev.* **8**, 2478–2490 (1994).
- Zelhof, A. C., Ghbeish, N., Tsai, C., Evans, R. M. & McKeown, M. A role for ultraspiracle, the *Drosophila* RXR, in morphogenetic furrow movement and photoreceptor cluster formation. *Development* **124**, 2499–2506 (1997).
- Brennan, C. A., Ashburner, M. & Moses, K. Ecdysone pathway is required for furrow progression in the developing *Drosophila* eye. *Development* **125**, 2653–2664 (1998).
- Cherbas, L., Hu, X., Zhimulev, I., Belyaeva, E. & Cherbas, P. EcR isoforms in *Drosophila*: testing tissue-specific requirements by targeted blockade and rescue. *Development* **130**, 271–284 (2003).
- Tsai, C. C., Kao, H. Y., Yao, T., McKeown, M. & Evans, R. M. SMRTER, a *Drosophila* nuclear receptor coregulator, reveals that EcR-mediated repression is critical for development. *Mol. Cell* **4**, 175–186 (1999).
- Hu, X., Cherbas, L. & Cherbas, P. Transcription activation by the ecdysone receptor: Identification of activation functions in EcR/USP. *Mol. Endocrinol.* **17**, 716–731 (2003).
- Ghbeish, N. & McKeown, M. Analyzing the repressive function of ultraspiracle, the *Drosophila* RXR, in *Drosophila* eye development. *Mech. Dev.* **111**, 89–98 (2002).
- Treisman, J. E. & Rubin, G. M. *wingless* inhibits morphogenetic furrow movement in the *Drosophila* eye disc. *Development* **121**, 3519–3527 (1995).
- Ma, C., Zhou, Y., Beachy, P. A. & Moses, K. The segment polarity gene hedgehog is required for progression of the morphogenetic furrow in the developing *Drosophila* eye. *Cell* **75**, 927–938 (1993).
- Heberlein, U., Wolff, T. & Rubin, G. M. The TGFβ homolog *dpp* and the segment polarity gene hedgehog are required for propagation of a morphogenetic wave in the *Drosophila* retina. *Cell* **75**, 913–926 (1993).
- Treisman, J. E., Luk, A., Rubin, G. M. & Heberlein, U. *eyelid* antagonizes *wingless* signaling during *Drosophila* development and has homology to the Bright family of DNA-binding proteins. *Genes Dev.* **11**, 1949–1962 (1997).
- Chanut, F., Luk, A. & Heberlein, U. A screen for dominant modifiers of *ro*^{Dom}, a mutation that disrupts morphogenetic furrow progression in *Drosophila*, identifies *groucho* and *hairless* as regulators of atonal expression. *Genetics* **156**, 1203–1217 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank C. Thummel, F. Kafatos, J. Treisman, M. Fujioka, T. Kornberg and Y. Zhang for antibodies, mutant stocks and mutant histones; and J. Kumar and F. R. Turner for the scanning electron microscopy images. This work was supported by a grant from the National Cancer Institute (to A.M.); an NIH award (to J.B.J.); a grant from the NSF (to P.C.); and an NIH training grant (to E.C.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.M. (A_Mazo@mail.jci.tju.edu).

Hedgehog signalling in the mouse requires intraflagellar transport proteins

Danwei Huangfu^{1,3}, Aimin Liu^{1,2}, Andrew S. Rakeman^{1,4}, Noel S. Murcia⁵, Lee Niswander^{1,2} & Kathryn V. Anderson¹

¹Developmental Biology Program, Sloan-Kettering Institute, 1275 York Avenue, New York, New York 10021, USA; ²Howard Hughes Medical Institute

³Neuroscience Program, and ⁴Molecular Cell and Developmental Biology Program; Weill Graduate School of Medical Sciences, Cornell University, 445 East 69th Street, New York, New York 10021, USA

⁵Department of Pediatrics, Rainbow Babies and Children's Hospital, Case Western Reserve University, 11100 Euclid Avenue, Cleveland, Ohio 44106, USA

Intraflagellar transport (IFT) proteins were first identified as essential factors for the growth and maintenance of flagella in the single-celled alga *Chlamydomonas reinhardtii*¹. In a screen for embryonic patterning mutations induced by ethylnitrosourea, here we identify two mouse mutants, *wimble* (*wim*) and *flexo* (*fxo*), that lack ventral neural cell types and show other phenotypes characteristic of defects in Sonic hedgehog signalling. Both mutations disrupt IFT proteins: the *wim* mutation is an allele of the previously uncharacterized mouse homologue of *IFT172*; and

fxo is a new hypomorphic allele of *polaris*, the mouse homologue of IFT88. Genetic analysis shows that Wim, Polaris and the IFT motor protein Kif3a are required for Hedgehog signalling at a step downstream of Patched1 (the Hedgehog receptor) and upstream of direct targets of Hedgehog signalling. Our data show that IFT machinery has an essential and vertebrate-specific role in Hedgehog signal transduction.

We identified *wim* and *fxo* as recessive mutations that caused abnormal embryonic morphology at mid-gestation. *wim* embryos arrested at embryonic day 10.5–11.5 (E10.5–11.5) with an open neural tube in the head that lacked the normal groove on the ventral midline (Fig. 1b)—a phenotype characteristic of mutants that lack Sonic hedgehog (Shh)-dependent ventral neural cell types^{2,3}. *fxo* mutant embryos had an unusually sharp angle of the mesencephalic flexure at E9.5–10.5 (Fig. 1c) and arrested at E12.5–13.5 with abnormal brain morphology and preaxial polydactyly—phenotypes that are also associated with abnormal Shh signalling.

We mapped *fxo* to a 2.0-megabase interval on chromosome 14 that included *polaris*, the homologue of *Chlamydomonas* IFT88 (Methods). Embryos homozygous for the null *polaris*^{tm1Rpw} allele lack nodal cilia, which causes randomized left–right asymmetry, and arrest at E10.5 with abnormal brain morphology⁴. We found that *fxo/polaris*^{tm1Rpw} transheterozygous embryos died at E12.5–13.5 with morphology similar to that of *fxo* homozygotes (data not shown). Sequence analysis showed that exon 16 was deleted in the *polaris* transcript in *fxo* mutants, causing an in-frame deletion of 29 amino acids including part of one of the ten predicted tetratricopeptide repeats (Supplementary Fig. 1a). Thus, *fxo* is a hypomorphic allele of *polaris*, which we hereafter designate *polaris*^{fxo}.

We mapped *wim* to a 484-kilobase (kb) interval on chromosome 5. We evaluated all genes in that interval as candidates, and the only mutation that we identified in the *wim* chromosome was a T to C mutation in the mouse homologue of IFT172 (Methods and D. Cole, personal communication). This mutation causes a leucine to proline substitution (Leu1564Pro) near the carboxy terminus of the protein, in a region that is highly conserved in homologous genes from *Caenorhabditis elegans* to humans (Supplementary Fig. 1b).

Chlamydomonas IFT172 and IFT88 are present in the same IFT protein complex, suggesting that mutations in the mouse homologues of these genes should cause similar phenotypes. We found that Wim, like Polaris, was required for left–right asymmetry: the polarity of heart looping was reversed in 40% (58 of 146) of E9.5 *wim* embryos (data not shown), and an early marker of left–right asymmetry, *nodal-lacZ* (ref. 5), showed abnormal bilateral expression in *wim* (Fig. 1e, f). Like *polaris*^{tm1Rpw} mutants, the node cells of *wim* embryos lacked all cilia (Fig. 1g–j), showing that Wim is required for the formation of nodal cilia and supporting the conclusion that the *wim* phenotype is caused by loss of IFT172 function.

wim and *polaris*^{tm1Rpw} embryos (Fig. 1d and data not shown) lacked the groove that is normally present at the ventral midline of the neural plate, a characteristic of mutants that lack ventral neural cell types^{2,3}. We therefore examined neural patterning in *polaris* and *wim* mutant embryos. We confirmed that *polaris*^{tm1Rpw} homozygous embryos do not express markers of the floor plate (ref. 4 and Fig. 2b), and found that other ventral cell types, including V3 interneurons and motor neurons, were also absent in *polaris*^{tm1Rpw} null mutants (Fig. 2f and data not shown). *polaris*^{fxo} embryos had a weaker neural phenotype: the floor plate was absent at most levels in the neural tube (Fig. 2c), whereas markers of V3 interneurons and motor neurons were expressed (Fig. 2g and data not shown).

The neural patterning phenotype of *wim* was indistinguishable from that of *polaris*^{tm1Rpw} embryos: markers of the floor plate, V3 interneuron progenitors and motor neurons were not expressed (Fig. 2d, h, and Supplementary Figs 2a–d and 3e–f). Embryos that carried the *wim* mutation in *trans* to *Hdh*^{df4J}, a large deletion that

uncovers the *wim* interval⁶, expressed neural markers in the same pattern as *wim* homozygous embryos (data not shown), indicating that *wim* is a null or nearly null allele. The loss of ventral neural cell types in *wim* and *polaris*^{tm1Rpw} was slightly less severe than that of *Shh* mutants, but very similar to that of double mutants lacking Gli2 and Gli3, the transcription factors that regulate Hedgehog (Hh) target gene expression⁷. For example, like *Gli2 Gli3* double mutants and unlike *Shh* mutants^{8,9}, Pax7 was expressed in its normal dorsal domain in *polaris*^{tm1Rpw} and *wim* (Fig. 2m–p) and some lateral cell fates (such as V2 interneurons) were specified (Supplementary Fig. 2e, f).

Mouse Kif3a, a subunit of the kinesin-II IFT motor, is also required for the formation of nodal cilia and normal left–right asymmetry^{10,11}. We found that E9.5 *Kif3a* mutant embryos did not specify a floor plate or V3 interneuron progenitors and expressed the lateral marker Pax6 at high levels across the ventral midline of the neural tube (Supplementary Fig. 4). Thus, three IFT proteins,

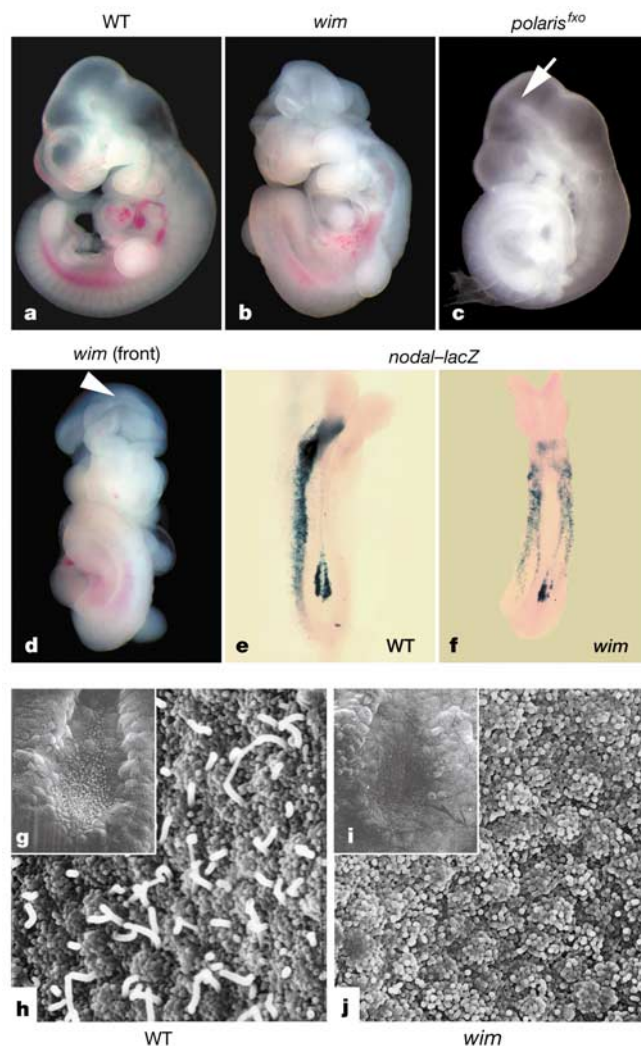


Figure 1 Phenotypes of *wim* and *fxo* mutants. **a–d**, In contrast to the E10.5 wild-type (WT) embryo (**a**), *wim* embryos have an open anterior neural tube (**b**) and *fxo* embryos (**c**) have a characteristically sharp angle at the mesencephalic flexure (arrow). Notably, *wim* embryos lack a groove on the ventral midline of the anterior neural tube (**d**, arrowhead). **e, f**, At E8.5, *nodal-lacZ* is normally expressed in the left lateral plate mesoderm (**e**); in five out of eight *wim* embryos, *nodal-lacZ* expression was observed bilaterally (**f**), and three other *wim* embryos showed the wild-type expression pattern. **g–j**, SEM analysis of embryonic node of E8.0 wild-type (**g, h**) and *wim* (**i, j**) embryos, at original magnification $\times 1,000$ (**g, i**) and $\times 6,600$ (**h, j**), shows that *wim* embryos lack nodal cilia.

Wim, Polaris and Kif3a, are all required for specification of ventral neural cell types.

The notochord is the initial source of the Shh ligand that specifies ventral cell types in the neural tube. The notochord is derived from the node^{8,9}, a structure affected by the IFT mutants^{4,10,11}. We found, however, that the notochord was specified normally and expressed wild-type amounts of Shh in E9.5 *wim*, *polaris* and *Kif3a* embryos (Supplementary Figs 3a–d and 4a, b, and data not shown), suggesting that the neural patterning defect is not the result of abnormal notochord development.

To test whether the IFT mutants lack ventral neural cell types because of reduced Shh signalling, we examined the expression of *Ptch1-lacZ*, a direct transcriptional target of Hh signalling¹². *Ptch1-lacZ* expression was greatly reduced in the neural tube of *wim*, *polaris*^{fox} and *Kif3a* mutants (Fig. 3a–h), indicating that the IFT mutants did not specify ventral cell types because they failed to activate Shh targets normally. *Ptch1-lacZ* was expressed in other sites of Hh signalling in wild-type embryos, including somites, the posterior limb bud and the gut (Fig. 3a, b). The expression of *Ptch1-lacZ* was greatly reduced at these sites in *wim*, *polaris*^{fox} and *Kif3a* mutants (Fig. 3c–h), indicating that these IFT genes are required for Hh signalling in all of these tissues.

To test whether the IFT proteins function within neural cells to

facilitate the normal response to Shh from the notochord, we examined double mutants with *Ptch1*. *Ptch1*, the receptor for Shh, is required to keep the signalling pathway 'off' in the absence of ligand^{9,12}. *Ptch1* mutants arrest at E9.0–9.5 with a distinctive morphology and express high levels of Shh target genes¹². In contrast to *Ptch1* mutants, which appear truncated anteriorly and fail to close the neural tube (ref. 12 and Fig. 3i), *wim Ptch1*, *polaris*^{fox} *Ptch1* and *Kif3a Ptch1* double mutants were morphologically indistinguishable from the respective *wim*, *polaris*^{fox} and *Kif3a* single mutants (Fig. 3). *Ptch1-lacZ* was highly expressed in most cells of *Ptch1* homozygous mutants (Fig. 3i, j). By contrast, *wim Ptch1*, *polaris*^{fox} *Ptch1* and *Kif3a Ptch1* double mutants expressed the same low levels of *Ptch1-lacZ* as *wim*, *polaris*^{fox} and *Kif3a* single mutants (Fig. 3k–p). Because the IFT mutations blocked the activation of the Shh pathway caused by loss of *Ptch1*, IFT proteins must be required for a step in the Shh signalling pathway downstream of *Ptch1* inside neural cells.

To define the step in the Shh signal transduction pathway affected by *wim* and *polaris* more precisely, we analysed neural patterning in additional double mutant embryos. *Rab23* (also called *open brain*) encodes a member of Rab family of vesicle transport proteins and

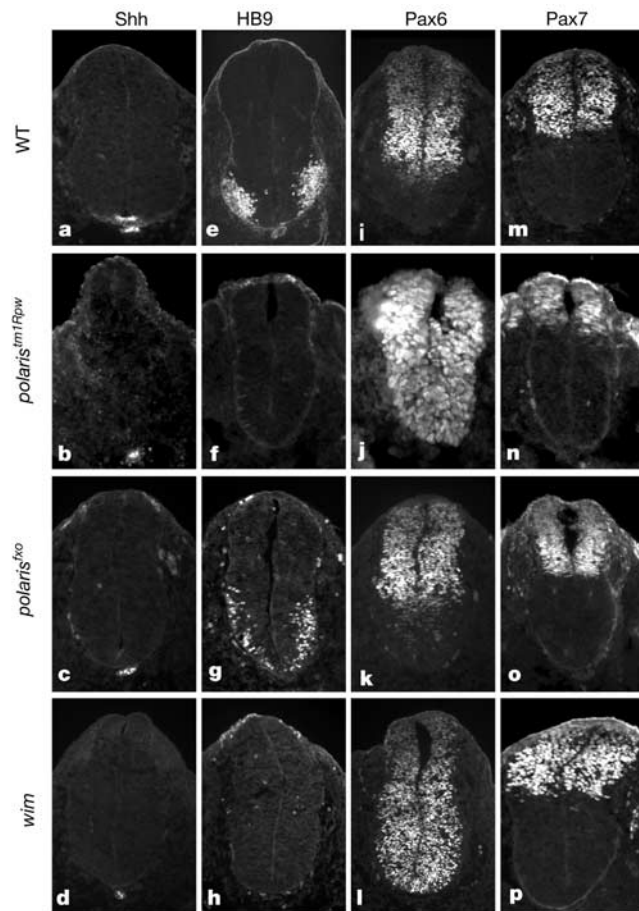


Figure 2 Dorso-ventral patterning in *polaris*^{tm1Rpw}, *polaris*^{fox} and *wim* mutant neural tubes. Immunohistochemical staining for Shh (a–d), HB9 (e–h), Pax6 (i–l) and Pax7 (m–p) in E10.5 wild-type (a, e, i, m), *polaris*^{tm1Rpw} (b, f, j, n), *polaris*^{fox} (c, g, k, o) and *wim* (d, h, l, p) embryos. *wim* and *polaris*^{tm1Rpw} mutants did not express markers for the floor plate (Shh) and motor neurons (HB9), although a few cells expressing HB9 were detected in some sections, and ventral neural cells instead expressed large amounts of the lateral marker Pax6. The same pattern of marker expression was seen at all rostro-caudal positions.

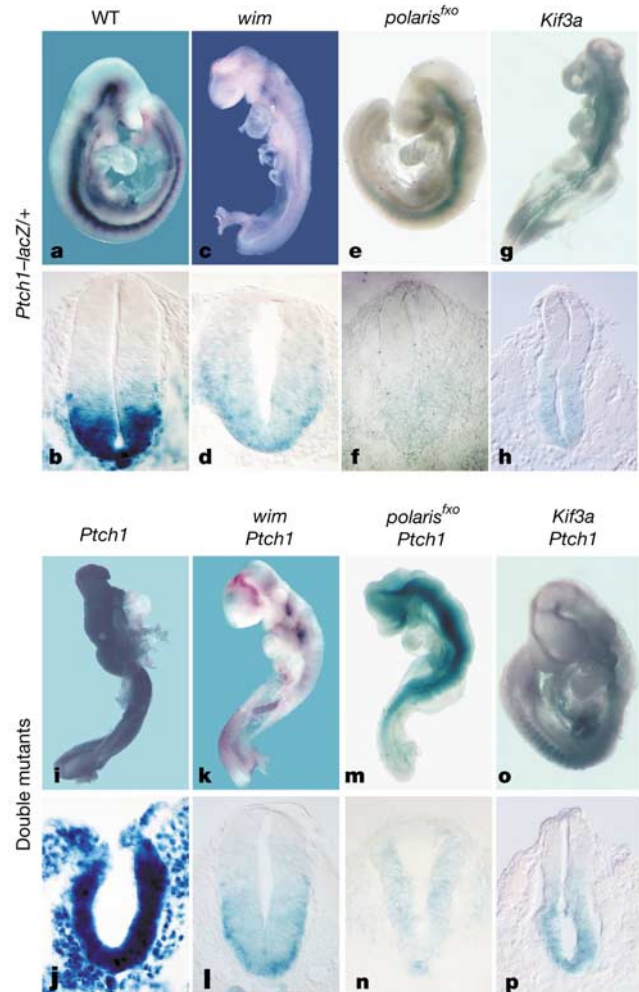


Figure 3 *Ptch1-lacZ* expression in single and double mutants. a–h, Expression of *Ptch1-lacZ* in embryos carrying one copy of the *Ptch1-lacZ* reporter. Shown are whole-mount (a, c, e, g) E9.5 embryos with corresponding sections beneath (b, d, f, h): wild-type (a, b), *wim* (c, d), *polaris*^{fox} (e, f) and *Kif3a* (g, h). i–p, Expression of *Ptch1-lacZ* in double mutant embryos. Shown are whole-mount embryos (i, k, m, o) with corresponding sections beneath (j, l, n, p): *Ptch1* homozygous mutant (i, j) compared with *wim Ptch1* (k, l), *polaris*^{fox} *Ptch1* (m, n) and *Kif3a Ptch1* (o, p) double mutants.

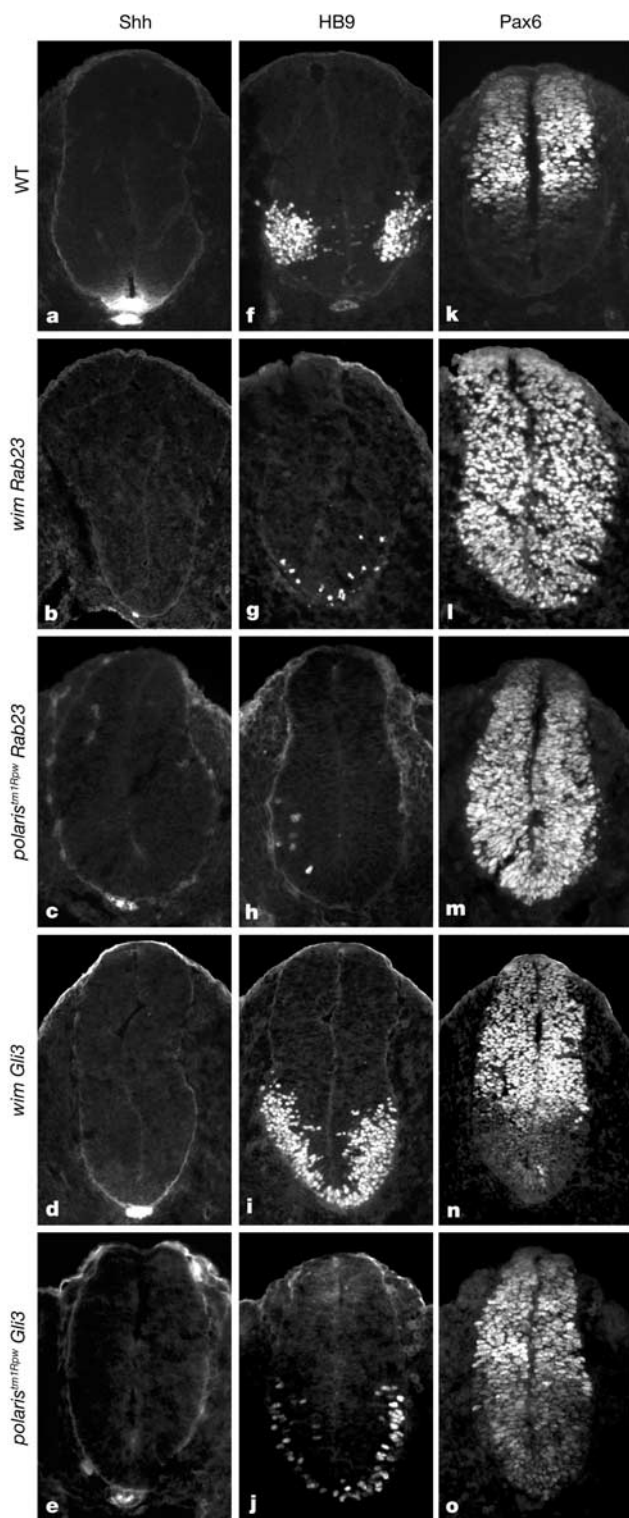


Figure 4 Dorso-ventral neural patterning in double mutant embryos. Shown is the expression of *Shh* (a–e), the motor neuron marker *HB9* (f–j) and *Pax6* (k–o) in the spinal cord of E10.5 wild-type (a, f, k), and *wim Rab23* (b, g, l), *polaris^{tm1Rpw} Rab23* (c, h, m), *wim Gli3* (d, i, n) and *polaris^{tm1Rpw} Gli3* (e, j, o) double mutants. Both *wim Rab23* and *polaris^{tm1Rpw} Rab23* double mutants lacked ventral neural cell types. By contrast, motor neurons were present and *Pax6* was expressed in its normal lateral domain in *wim Gli3* and *polaris^{tm1Rpw} Gli3* double mutants. More motor neurons were specified in caudal than in rostral regions of the *wim Gli3* and *polaris^{tm1Rpw} Gli3* neural tube; similar results have been observed in *Shh Gli3* double mutants¹⁶.

negatively regulates the Hh pathway at a point downstream of *Ptch1* and *Smoothed* (*Smo*; refs 13, 14, and J. Eggenschwiler and K.V.A., unpublished data). In contrast to the expansion of ventral cell fates in *Rab23* mutants^{13,14}, *wim Rab23* and *polaris^{tm1Rpw} Rab23* double mutants failed to specify ventral neural cell types (Fig. 4b, c, g, h) like *wim* and *polaris^{tm1Rpw}* single mutants (Fig. 2b, d, f, h). Thus, *wim* and *polaris* are required for signalling in the *Shh* pathway at a step downstream of *Ptch1*, *Smo* and *Rab23*. The *Gli3* protein can be proteolytically processed into a repressor form that negatively regulates Hh target genes^{9,15}. Motor neurons, which were not specified in *wim* and *polaris^{tm1Rpw}* mutants (Fig. 2f, h), were rescued in *wim Gli3* and *polaris^{tm1Rpw} Gli3* double mutants (Fig. 4i, j). Similar results have been seen with *Shh*: motor neurons that are absent in *Shh* single mutants are specified in *Shh Gli3* double mutant neural tubes¹⁶. A simple interpretation of these results is that the wild-type *Wim* and *Polaris* antagonize production or activity of the *Gli3* repressor.

Two types of model could explain how IFT proteins are required for Hh signalling pathway: one is based on a function in cilia and the second is based on two separate functions in cilia formation and intracellular transport. In the first scheme, IFT proteins are required only for the production of cilia, and cilia produce a signal that is required for a step in the Hh signalling pathway downstream of *Ptch1*, *Smo* and *Rab23*. Most vertebrate cells, including neuronal precursors, have non-motile primary cilia¹⁷ and there is evidence that signals originating in cilia can affect cellular physiology.

In the second scheme, IFT proteins in the developing mouse embryo have two separate functions: one in cilia formation and a second in intracellular transport. Two other trafficking proteins that function in the Hh pathway have been identified: the *Drosophila* kinesin-like protein *Costal2* (*Cos2*) and mouse *Rab23*, both of which are negative regulators that function downstream of *Smo* and upstream of the *Ci/Gli* transcription factors^{13,18}. It is possible that *Rab23* (and perhaps *Cos2*) directs *Gli* proteins to a location where proteolytic processing creates the *Gli* repressor forms and that IFT proteins traffic *Gli* proteins away from the site of processing.

Whereas most vertebrate cells are ciliated¹⁷, cilia in invertebrates are largely confined to neurons and the phenotypes caused by IFT mutations in invertebrates specifically affect ciliated sensory neurons and axonal transport^{11,19}. Null mutations in the *Drosophila* *Kif3a* and *IFT88* homologues, *Klp64D* and *nompB*, cause neuronal phenotypes but do not seem to disrupt the Hh signalling^{19,20}. Thus, although IFT proteins have an evolutionarily conserved role in the assembly of cilia and flagella, we propose that the requirement for IFT proteins in the Hh signal transduction pathway is specific to vertebrates. Because ciliary function and the Hh signalling pathway are important in many different human pathologies, including Kartagener syndrome, pulmonary dysfunction, heterotaxia, polycystic kidney disease, retinal degeneration, infertility, hydrocephalus, holoprosencephaly, polydactyly and cancer^{1,21,22}, understanding the relationship between mammalian IFTs and Hh signalling will be important for understanding the biological bases of these human conditions. □

Methods

Mouse strains and genetic mapping

We identified *wim* and *fxo* in a screen for recessive ethylnitrosourea-induced mutations that cause morphological abnormalities at E9.5 (refs 3, 23). Other mutations used were *polaris^{tm1Rpw}* (ref. 4); *nodal^{tm1Rob}*, with a *lacZ* gene insertion in the *nodal* gene²; *Hdh^{d4f}* deletion²; *Kif3a^{tm1Maz}* (ref. 11; The Jackson Laboratory); *Ptch1^{tm1Mps}*, a targeted null allele of *Ptch1* with insertion of the *lacZ* gene¹²; *Rab23^{apb2}*, a nonsense allele recovered in our ENU screen^{13,14}; and *Gli3^{XT-1}* (The Jackson Laboratory), an intragenic deletion allele²⁴. Analysis of *wim* and *fxo* was done on a C3H congenic background; *polaris^{tm1Rpw}*, *Kif3a*, *Ptch1*, *Rab23* and *Gli3* were crossed onto C3H for at least two generations before analysis. All double mutants were obtained from intercrosses between double heterozygous animals. Littermates were used as wild-type controls. We genotyped mutant alleles as described^{4,5,11,12,14,24}.

Genetic mapping was done as described^{3,13}. Massachusetts Institute of Technology (MIT) simple sequence length polymorphism (SSLP) markers were used for initial

mapping. Additional polymorphic DNA markers based on repeat sequences were generated for high-resolution mapping (see <http://mouse.ski.mskcc.org/>).

Molecular identification of *fxo* and *wim*

In a mapping cross of 796 opportunities for recombination, *fxo* was mapped between D14Mit84 and D7Mit111. Products from polymerase chain reaction with reverse transcription (RT-PCR) that were amplified from E10.5 *fxo* embryos lacked the 87-base-pair exon 16 of the *polaris* transcript. Analysis of the genomic sequence surrounding exon 16 identified a T to C mutation in the second nucleotide of intron 16 (downstream of exon 16), which was associated with the skipping of exon 16.

In a mapping cross of 2,600 opportunities for recombination, *wim* was mapped to a 484-kb interval between D5Ski122 and D5Ski114 (see <http://mouse.ski.mskcc.org/>). The *wim* interval contained 7 known and 14 predicted genes, according to the current genome assembly. We ruled out four genes (*Slc30a3*, *Ucn*, *Mpv17* and *Gckr*) as candidates because mice lacking these genes survive to adulthood^{25–28}. Three genes were not expressed in mid-gestation embryos on the basis of RT-PCR analysis of E10.5 RNA. Four predicted genes were not sequenced because their predicted function seemed completely unrelated to Hh signalling or neural tube patterning. The complete coding region of the remaining ten candidate genes was sequenced from RT-PCR products amplified from E10.5 *wim* (six independent embryos sequenced) and parental (C57BL/6) embryos. The only sequence change identified was a T to C substitution in a predicted gene corresponding to five Ensembl predicted transcripts (ENSMUST000000041444, ENSMUST000000041565, ENSMUST000000041704, ENSMUST000000031036 and ENSMUST000000041963) homologous to the full-length rat *SLB* RNA²⁹. The change was confirmed in genomic DNA from *wim* embryos and is not present in expressed sequence tags from other mouse strains. This gene is expressed broadly at E9.5 and E10.5 and at high levels in the node of E7.5 embryos (data not shown).

We identified structural motifs in Polaris and Wim by SMART (<http://smart.embl-heidelberg.de>) and Pfam (<http://www.sanger.ac.uk/Software/Pfam/>) analysis.

Phenotypic analysis

In situ hybridization and X-Gal staining on whole-mount embryos and immunofluorescence on frozen sections were done as described¹⁴. For scanning electron microscopy (SEM), we collected embryos at E8.0, fixed them in 2.5% glutaraldehyde in 0.1 M Sorenson's phosphate for >24 h, and dehydrated them through an ethanol series. Samples were processed and observed according to standard procedures³⁰.

Received 15 May; accepted 3 September 2003; doi:10.1038/nature02061.

- Rosenbaum, J. L. & Witman, G. B. Intraflagellar transport. *Nature Rev. Mol. Cell Biol.* **3**, 813–825 (2002).
- Ma, Y. *et al.* Hedgehog-mediated patterning of the mammalian embryo requires transporter-like function of Dispatched. *Cell* **111**, 63–75 (2002).
- Caspar, T. *et al.* Mouse *Dispatched* homolog 1 is required for long-range, but not juxtacrine, Hh signaling. *Curr. Biol.* **12**, 1628–1632 (2002).
- Murcia, N. S. *et al.* The *Oak Ridge polycystic kidney (ork)* disease gene is required for left–right axis determination. *Development* **127**, 2347–2355 (2000).
- Collignon, J., Varlet, I. & Robertson, E. J. Relationship between asymmetric nodal expression and the direction of embryonic turning. *Nature* **381**, 155–158 (1996).
- Schimenti, J. C. *et al.* Interdigitated deletion complexes on mouse chromosome 5 induced by irradiation of embryonic stem cells. *Genome Res.* **10**, 1043–1050 (2000).
- Motoyama, J. *et al.* Differential requirement for Gli2 and Gli3 in ventral neural cell fate specification. *Dev. Biol.* **259**, 150–161 (2003).
- Chiang, C. *et al.* Cyclopia and defective axial patterning in mice lacking *Sonic hedgehog* gene function. *Nature* **383**, 407–413 (1996).
- Litingtung, Y. & Chiang, C. Control of Shh activity and signaling in the neural tube. *Dev. Dyn.* **219**, 143–154 (2000).
- Takeda, S. *et al.* Left-right asymmetry and kinesin superfamily protein KIF3A: new insights in determination of laterality and mesoderm induction by *kif3A*^{−/−} mice analysis. *J. Cell Biol.* **145**, 825–836 (1999).
- Marszałek, J. R., Ruiz-Lozano, P., Roberts, E., Chien, K. R. & Goldstein, L. S. *Situs inversus* and embryonic ciliary morphogenesis defects in mouse mutants lacking the KIF3A subunit of kinesin-II. *Proc. Natl Acad. Sci. USA* **96**, 5043–5048 (1999).
- Goodrich, L. V., Milenkovic, L., Higgins, K. M. & Scott, M. P. Altered neural cell fates and medulloblastoma in mouse *patched* mutants. *Science* **277**, 1109–1113 (1997).
- Eggenchwil, J. T., Espinoza, E. & Anderson, K. V. Rab23 is an essential negative regulator of the mouse *Sonic hedgehog* signalling pathway. *Nature* **412**, 194–198 (2001).
- Eggenchwil, J. T. & Anderson, K. V. Dorsal and lateral fates in the mouse neural tube require the cell-autonomous activity of the *open brain* gene. *Dev. Biol.* **227**, 648–660 (2000).
- Wang, B., Fallon, J. F. & Beachy, P. A. Hedgehog-regulated processing of Gli3 produces an anterior/posterior repressor gradient in the developing vertebrate limb. *Cell* **100**, 423–434 (2000).
- Litingtung, Y. & Chiang, C. Specification of ventral neuron types is mediated by an antagonistic interaction between Shh and Gli3. *Nature Neurosci.* **3**, 979–985 (2000).
- Wheatley, D. N., Wang, A. M. & Strugnell, G. E. Expression of primary cilia in mammalian cells. *Cell Biol. Int.* **20**, 73–81 (1996).
- Sisson, J. C., Ho, K. S., Suyama, K. & Scott, M. P. Costal2, a novel kinesin-related protein in the Hedgehog signaling pathway. *Cell* **90**, 235–245 (1997).
- Ray, K. *et al.* Kinesin-II is required for axonal transport of choline acetyltransferase in *Drosophila*. *J. Cell Biol.* **147**, 507–518 (1999).
- Han, Y., Kwok, B. H. & Kernan, M. J. Intraflagellar transport is required in *Drosophila* to differentiate sensory cilia but not sperm. *Curr. Biol.* **13**, 1679–1686 (2003).
- Bale, A. E. Hedgehog signaling and human disease. *Annu. Rev. Genom. Hum. Genet.* **3**, 47–65 (2002).
- Sloboda, R. D. A healthy understanding of intraflagellar transport. *Cell Motil. Cytoskeleton* **52**, 1–8 (2002).
- Kasarskis, A., Manova, K. & Anderson, K. V. A phenotype-based screen for embryonic lethal mutations in the mouse. *Proc. Natl Acad. Sci. USA* **95**, 7485–7490 (1998).

- Hui, C. C. & Joyner, A. L. A mouse model of Greig cephalopolysyndactyly syndrome: the *extra-toes*¹ mutation contains an intragenic deletion of the *Gli3* gene. *Nature Genet.* **3**, 241–246 (1993).
- Cole, T. B., Wenzel, H. J., Kafer, K. E., Schwartzkroin, P. A. & Palmiter, R. D. Elimination of zinc from synaptic vesicles in the intact mouse brain by disruption of the *ZnT3* gene. *Proc. Natl Acad. Sci. USA* **96**, 1716–1721 (1999).
- Vetter, D. E. *et al.* Urocortin-deficient mice show hearing impairment and increased anxiety-like behavior. *Nature Genet.* **31**, 363–369 (2002).
- Weihert, H., Noda, T., Gray, D. A., Sharpe, A. H. & Jaenisch, R. Transgenic mouse model of kidney disease: insertional inactivation of ubiquitously expressed gene leads to nephrotic syndrome. *Cell* **62**, 425–434 (1990).
- Farrelly, D. *et al.* Mice mutant for glucokinase regulatory protein exhibit decreased liver glucokinase: a sequestration mechanism in metabolic regulation. *Proc. Natl Acad. Sci. USA* **96**, 14511–14516 (1999).
- Howard, P. W. & Maurer, R. A. Identification of a conserved protein that interacts with specific LIM homeodomain transcription factors. *J. Biol. Chem.* **275**, 13336–13342 (2000).
- Sulik, K. *et al.* Morphogenesis of the murine node and notochordal plate. *Dev. Dyn.* **201**, 260–278 (1994).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank K. Maxwell and T. Caspar for initial experiments with *fxo*; N. Lampen for assistance with SEM; J. Eggenchwil, D. Cole and M. Kernan for sharing unpublished data; C. Sander and B. Reva for discussions about Wim and Polaris protein structures; T. Bestor, T. Caspar, J. Eggenchwil, M. García-García and J. Lee for comments on the manuscript; E. Robertson, M. Scott and J. Schimenti for mice; and C. Cepko for the Chx10 antibody. Monoclonal antibodies were obtained from the Developmental Studies Hybridoma Bank, which was developed under the auspices of the National Institute of Child Health and Human Development and is maintained by The University of Iowa, Department of Biological Sciences. Genome sequence analysis used Ensembl and the Celera Discovery System and associated databases, made possible in part by the AMDeC Foundation. This work was supported by NIH grants to K.V.A. and the Lita Annenberg Hazen Foundation. L.N. is an Investigator of the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.V.A. (k-anderson@ski.mskcc.org). The *wim* complementary DNA sequence is deposited in GenBank under accession number AY339616.

Degradation of Cdc25A by β -TrCP during S phase and in response to DNA damage

Luca Busino^{1*}, Maddalena Donzelli^{1*}, Massimo Chiesa¹, Daniele Guardavaccaro², Dvora Ganoth³, N. Valerio Dorrello², Avram Hershko³, Michele Pagano² & Giulio F. Draetta¹

¹European Institute of Oncology, 435 Via Ripamonti, 20141 Milan, Italy

²Department of Pathology, MSB 599, New York University School of Medicine and NYU Cancer Institute, 550 First Avenue, New York, New York 10016, USA

³Unit of Biochemistry, B. Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa 31096, Israel

*These authors equally contributed to this work

The Cdc25A phosphatase is essential for cell-cycle progression because of its function in dephosphorylating cyclin-dependent kinases. In response to DNA damage or stalled replication, the ATM and ATR protein kinases activate the checkpoint kinases Chk1 and Chk2, which leads to hyperphosphorylation of Cdc25A^{1–3}. These events stimulate the ubiquitin-mediated proteolysis of Cdc25A^{1,4,5} and contribute to delaying cell-cycle progression, thereby preventing genomic instability^{1–7}. Here we report that β -TrCP is the F-box protein that targets phosphorylated Cdc25A for degradation by the Skp1/Cul1/F-box protein complex. Downregulation of β -TrCP1 and β -TrCP2 expression by short interfering RNAs causes an accumulation of Cdc25A in cells progressing through S phase and prevents the degradation of Cdc25A induced by ionizing radiation, indicating that β -TrCP may function in the intra-S-phase checkpoint. Consistent with

this hypothesis, suppression of β -TrCP expression results in radioresistant DNA synthesis in response to DNA damage—a phenotype indicative of a defect in the intra-S-phase checkpoint that is associated with an inability to regulate Cdc25A properly. Our results show that β -TrCP has a crucial role in mediating the response to DNA damage through Cdc25A degradation.

We previously showed that the ubiquitin-mediated degradation of Cdc25A at the exit of mitosis is mediated by the anaphase promoting complex/cyclosome (APC/C^{Cdh1}) ubiquitin ligase through recognition of a specific KEN box sequence⁸. We also showed that a Cdc25A KEN mutant is resistant to APC-mediated ubiquitination but remains short-lived in interphase cells like the wild-type protein and is still degraded in response to ionizing radiation, thereby suggesting that Cdc25A may be targeted for degradation by a dual mechanism⁸. Furthermore, degradation of Cdc25A both in cycling cells and in response to DNA damage depends on phosphorylation^{1–4}, which is also a requirement for the efficient recruitment of target proteins to Skp1/Cullin/F-box (SCF) ubiquitin ligases by F-box proteins^{9,10}, as reported for other cell-cycle regulators (reviewed in ref. 11).

We previously established that Cdc25A is in complex with Cul1 and Skp1 (ref. 8), which prompted us to examine the possible involvement of SCF in the degradation of Cdc25A. To achieve this, we screened several human F-box proteins^{12,13} for interaction with Cdc25A. We observed that β -TrCP1 and β -TrCP2 (also known as Fbw1a and Fbw1b, respectively) were the only F-box proteins among those tested, namely Fbw2, Fbw4, Fbw5 and Fbw7, that interacted with Cdc25A *in vivo* (Fig. 1a and data not shown for Skp2 and Fbw3). As a control, we assessed that the Cul1 subunit was bound to all of the F-box proteins tested. In a parallel experiment, we showed that overexpressed Cdc25A coprecipitates with β -TrCP1 and β -TrCP2 (hereafter β -TrCP1/2), but not with Skp2 and Fbw7 (Fig. 1b). Notably, the Cdc25A that co-precipitated with β -TrCP1/2 was hyperphosphorylated, because treatment of the immunocomplexes with λ -phosphatase generated a faster migrating band of Cdc25A (Fig. 1c).

The Cdc25A amino acid sequence contains a motif, DSGXXXXS (DSG(X)₄S), which is similar to the DSGXXS or DSGXXXX motifs present in known protein substrates of the SCF ^{β -TrCP} complex (refs 14–17 and Fig. 2a). Phosphorylation at both serine residues is required both for β -TrCP binding and for subsequent substrate degradation^{15,18}. To assess the contribution of this motif to the interaction, we mutated Ser 82 and Ser 88 to alanine, either alone to generate a Cdc25A^{DSG2X} mutant, or in addition to the preceding Ser 79 to produce a Cdc25A^{DSG3X} mutant (Fig. 2a). We then tested whether these mutated proteins would interact with β -TrCP in transfected cells.

Replacement of the two DSG(X)₄S serines with alanine was sufficient to abolish Cdc25A interaction with β -TrCP1 (Fig. 2b). This interaction is probably dependent on phosphorylation on the DSG(X)₄S serine residues. Indeed, a Cdc25A-derived peptide containing phosphoserine residues at positions Ser 82 and Ser 88 (Fig. 2a) associated with *in vitro* translated β -TrCP1/2, but not with other F-box proteins tested, whereas the unphosphorylated peptide did not associate at all (Fig. 2c). We also assessed whether a failure to interact with β -TrCP would affect the stability of Cdc25A. We determined the half-life of Cdc25A^{DSG2X} and Cdc25A^{DSG3X} and found that the mutants were considerably more stable than wild-type Cdc25A (Fig. 2d).

To investigate whether β -TrCP stimulates ubiquitination of Cdc25A, we carried out an *in vitro* ubiquitination assay using partially reconstituted HeLa cell extracts. Addition of recombinant purified β -TrCP1, but not two other F-box proteins (Skp2 and Fbw7), stimulated ubiquitination of Cdc25A (Fig. 2e). Ubiquitination by β -TrCP also required protein phosphorylation, because the reaction was inhibited by replacing ATP with AMP-PNP, an analogue that allows ubiquitin adenylation by E1 ligase but cannot

function as a protein kinase substrate (Fig. 2f). As a control, AMP-PNP facilitated APC-dependent ubiquitination, which does not require substrate phosphorylation (data not shown). These data support the hypothesis that Cdc25A phosphorylation is required for β -TrCP-mediated ubiquitination. Notably, β -TrCP1-mediated ubiquitination of Cdc25A^{DSG3X} was greatly affected as compared with wild-type Cdc25A (Fig. 2g). Together, these data indicate that phosphorylation on Ser 82 and Ser 88 is required for efficient recruitment of Cdc25A by β -TrCP and its subsequent degradation.

We used a series of siRNA experiments to assess further the role of β -TrCP in controlling the subcellular abundance of Cdc25A. Asynchronously growing cells were transiently transfected with an siRNA oligonucleotide targeting both β -TrCP1 and β -TrCP2 genes and analysed for steady-state amounts of Cdc25A. Efficient silencing of β -TrCP1/2 caused a considerable accumulation of Cdc25A as compared with mock-transfected cells (Fig. 3a). No substantial changes were observed in Cdc25A messenger RNA (Fig. 3a).

We then tested the effect of silencing β -TrCP gene expression at various stages of the cell cycle. Cells that were mock-transfected or transfected with β -TrCP1/2 siRNA were synchronized by double-

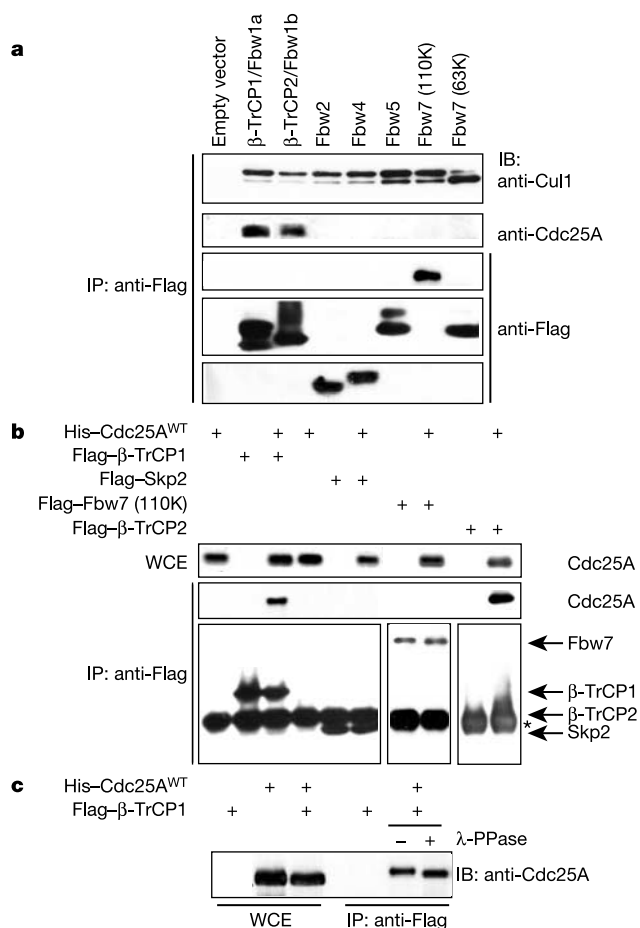


Figure 1 Cdc25A interacts with β -TrCP1 and β -TrCP2 *in vivo*. **a**, HeLa cells were transfected with the indicated Flag-tagged F-box protein constructs. F-box proteins were immunoprecipitated from extracts with an anti-Flag resin and immunocomplexes were blotted with antibodies specific for Cul1, Cdc25A and Flag. **b**, HeLa cells were co-transfected with a vector expressing His-Cdc25A plus the indicated Flag-tagged F-box protein constructs. Immunocomplexes were analysed as in **a**. WCE, whole-cell extract. Asterisk indicates the position of the IgG heavy chain (β -TrCP2 protein overlaps with IgG). **c**, Cdc25A bound to β -TrCP proteins is phosphorylated. Immunocomplexes obtained by β -TrCP1 immunoprecipitation were treated with λ -phosphatase and analysed for Cdc25A.

thymidine treatment, released from G1/S arrest in the presence of nocodazole and analysed over time for Cdc25A. As they progressed through S phase, cells treated with β -TrCP1/2 siRNA showed a substantial accumulation of Cdc25A as compared to mock-transfected cells, as well as a failure to degrade the APC inhibitor Emi1 (also known as Fbx5), a target of β -TrCP in early mitosis (refs 19, 20 and Fig. 3b).

To analyse the kinetics of Cdc25A expression at mitotic exit and in G1 phase, mock-transfected or cells transfected with β -TrCP1/2

siRNA were synchronized by nocodazole treatment, released from the mitotic block and analysed over time for Cdc25A. Cells transfected with β -TrCP1/2 siRNA had substantially more Cdc25A at mitosis than did mock-transfected cells, but similar to control cells they proceeded normally into G1 phase and degraded both Cdc25A and cyclin B1, albeit with slower kinetics (Fig. 3c). This behaviour is probably caused by an increase in Emi1 in cells transfected with β -TrCP siRNA^{19,20}, resulting in an indirect up-regulation of Cdc25A through inhibition of Cdh1 at the exit of mitosis²¹.

We directly compared the amounts of Cdc25A in mock-transfected and cells transfected with β -TrCP1/2 and Emi1 siRNA at different time points (Fig. 3b, c, right). Elimination of β -TrCP caused an accumulation of Cdc25A in all cell-cycle phases apart from G1 (Fig. 3c, T6). As a control, we verified that elimination of

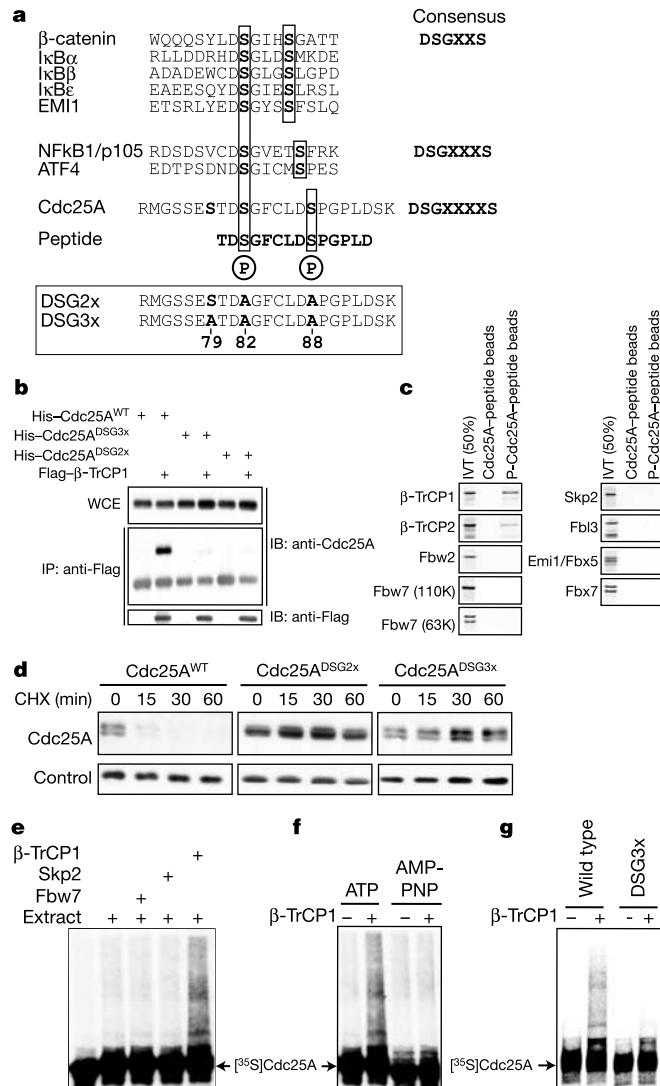


Figure 2 Interaction with β -TrCP protein through a phosphorylated DSG motif is required for Cdc25A degradation and polyubiquitination. **a**, Alignment of DSG motifs identified in known β -TrCP substrates. Serine-to-alanine mutations in the DSG motif of Cdc25A are boxed. **b**, Vectors expressing His-tagged Cdc25A^{DSG2x} and Cdc25A^{DSG3x} were coexpressed with Flag- β -TrCP1 in HeLa cells, and anti-Flag immunocomplexes were blotted for Cdc25A. **c**, Immobilized Cdc25A-derived peptides, with or without phosphorylation on Ser 82 and Ser 88 (**a**), were incubated with [³⁵S]methionine-labelled *in vitro* translated F-box proteins (IVT) and analysed by autoradiography. **d**, HeLa cells were transfected with the indicated Flag-tagged constructs and analysed for Cdc25A expression after cycloheximide (CHX) treatment. **e**, [³⁵S]methionine-labelled *in vitro* translated Cdc25A was incubated with HeLa cell extract enriched with the indicated Skp1-F-box protein complexes and analysed by autoradiography. **f**, The ATP analogue AMP-PNP was used in the ubiquitin ligation reaction. **g**, β -TrCP-mediated ubiquitination of wild-type Cdc25A and the Cdc25A^{DSG3x} mutant.

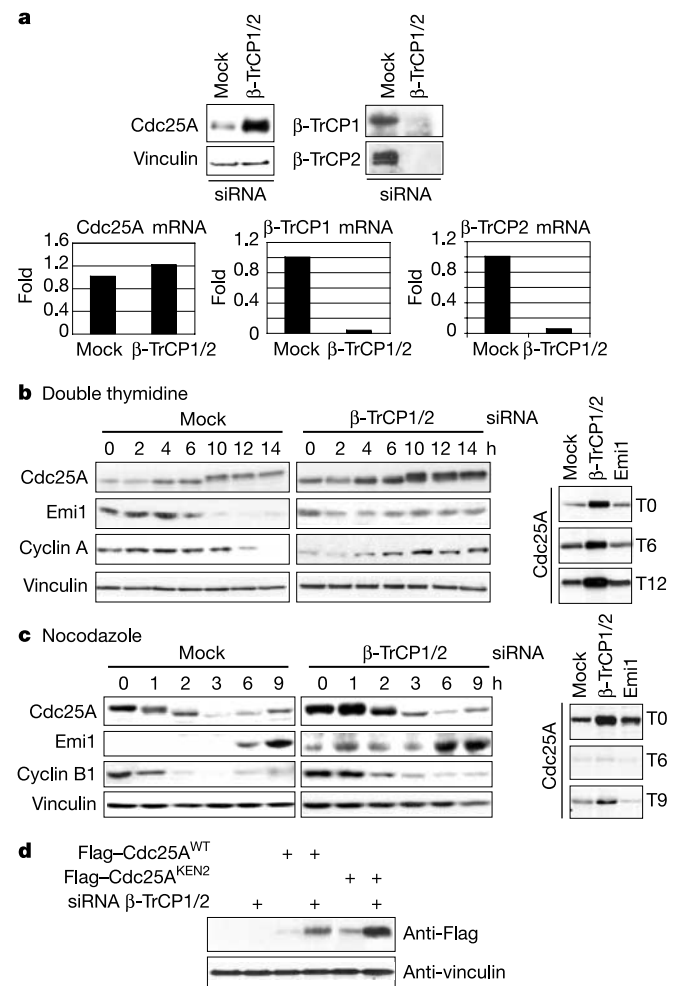


Figure 3 β -TrCP controls Cdc25A abundance during progression of S phase. **a**, HeLa cells were mock-transfected or transfected with β -TrCP1 and β -TrCP2 siRNA oligonucleotides. Expression of Cdc25A and β -TrCP1/2 (analysed by immunoblotting and quantitative PCR) is shown. **b**, Cells transfected as indicated were synchronized by double-thymidine block and released in nocodazole-containing medium. Cells were collected at the indicated time points, lysed and immunoblotted for Cdc25A, Emi1 and cyclin A. Samples collected at the T0, T6 and T12 time points were aligned for a direct comparison of Cdc25A expression. **c**, Cells transfected as indicated were synchronized by nocodazole treatment and released in drug-free medium. Cells were analysed for Cdc25A, Emi1 and cyclin B1 as in **b**. **d**, Cells transfected with a vector expressing Flag-tagged wild-type Cdc25A or a Cdc25A^{KEN2} mutant were subjected to RNA interference and analysed for overexpression of Cdc25A.

Emi1 did not affect the expression of Cdc25A protein. We also observed that the Cdc25A^{KEN2} mutant, which is not degraded by APC/C^{Cdh1}, also accumulated in cells transfected with β -TrCP1/2 siRNA (Fig. 3d). Together, these data indicate that β -TrCP-mediated degradation of Cdc25A may occur throughout S and G2, and that this event is independent of the release of the Emi1-mediated inhibition of Cdh1.

To examine whether β -TrCP proteins are involved in the ubiquitin-mediated degradation of Cdc25A in response to DNA damage, cells transfected with β -TrCP1/2 siRNA were treated with ionizing radiation and analysed for Cdc25A abundance. Cells with reduced β -TrCP1/2 showed an increase in Cdc25A as compared with mock-transfected and Cdh1-depleted cells, which were used as a negative control. Notably, ionizing radiation resulted in an accumulation of slow-migrating, hyperphosphorylated Cdc25A species (Fig. 4a, b). Furthermore, the half-life of Cdc25A in S-phase-synchronized and

β -TrCP1/2-depleted cells was extended and remained unmodified in cells treated with ionizing radiation as compared with untreated cells (Fig. 4b).

To establish a role for β -TrCP in the Cdc25A-mediated DNA damage response, we examined whether interfering with β -TrCP expression would result in a defect in the temporal inhibition of DNA synthesis. Radioresistant DNA synthesis (RDS), a phenotype indicative of a defective intra-S-phase checkpoint, occurs in cells deficient for ATM and Chk2, two upstream negative effectors of Cdc25A abundance^{1,22}. We checked for the integrity of the intra-S-phase checkpoint in cells transfected with β -TrCP siRNA by assessing their rate of DNA synthesis after treatment with ionizing radiation. Whereas mock-transfected cells showed an inhibition of DNA synthesis of about 40–50% after ionizing radiation, β -TrCP1/2-depleted cells showed an inhibition of roughly 20%, consistent with an RDS phenotype (Fig. 4c). This effect was dependent on an accumulation of Cdc25A caused by β -TrCP inhibition, given that cells depleted for both β -TrCP and Cdc25A rescued the RDS phenotype.

Using a polyclonal antibody raised against the phosphorylated DSG motif of Cdc25A (anti-phosphoS82/S88) and phosphopeptide mapping, we detected Cdc25A phosphorylation at the DSG motif in cycling cells (Supplementary Figs 1 and 2). DSG phosphorylation was stimulated on treatment with ionizing radiation (Fig. 4d). It therefore seems that Cdc25A is phosphorylated at the DSG motif in cycling cells and that this process is enhanced on DNA damage. Indeed, expression of the Cdc25A^{DSG2x} mutant hampered Cdc25A degradation on treatment with ionizing radiation (Fig. 4e).

Cdc25A is phosphorylated on specific serine residues by the Chk1 protein kinase in cycling cells and by both Chk1 and Chk2 in checkpoint activated cells³, and disruption of the Chk1/Cdc25A pathway abrogates S and G2 checkpoints induced by ionizing radiation^{2,3}. Combined mutation of these residues confers stability to the protein, suggesting that phosphorylation at several sites contributes to Cdc25A degradation *in vivo*³. These findings might indicate that, as in the recognition of Sic1 by Fbw7 (ref. 23), multiple phosphorylation events on the target protein enhance its interaction with the F-box protein and stimulate its polyubiquitination²⁴. For Cdc25A, we found that whereas the Cdc25A^{DSG2x} mutant failed to form complexes *in vivo* with β -TrCP1, Skp1 or Cul1, none of the Chk1/2 phosphorylation site mutants was impaired in binding these SCF components (Fig. 4f). The finding that the DSG motif is required for β -TrCP binding (Fig. 4f) and for β -TrCP-dependent ubiquitination (Fig. 2g) implies that efficient Cdc25A recognition by β -TrCP minimally requires two phosphorylation sites in Cdc25A. This is in agreement with the three-dimensional structure of a β -TrCP1–Skp1– β -catenin complex²⁵. Phosphorylation on serine residues other than those in the DSG motif³ might stimulate the degradation of Cdc25A by facilitating an interaction with additional components of the SCF complex, or by enhancing the ability of SCF to catalyse polyubiquitination.

We have shown a requirement for β -TrCP proteins in regulating Cdc25A ubiquitination in S phase and on treatment with ionizing radiation. Notably, a gain-of-function mutation in the *cdc25* gene of *Caenorhabditis elegans* results in deregulated hyperproliferation of intestinal cells²⁶. The encoded mutated protein carries a serine-to-phenylalanine substitution in a putative DSG consensus. Given the fact that mutations in ATM and ATR, which are upstream negative effectors of Cdc25A, are common in cancer^{22,27}, an assessment of the occurrence of activating mutations in the Cdc25A gene and/or inactivating mutations in the β -TrCP genes in normal and tumour tissues should be warranted. □

Methods

Cell culture, synchronization and transfection

Cell culture, cell synchronization and transfection protocols were done as described⁸.

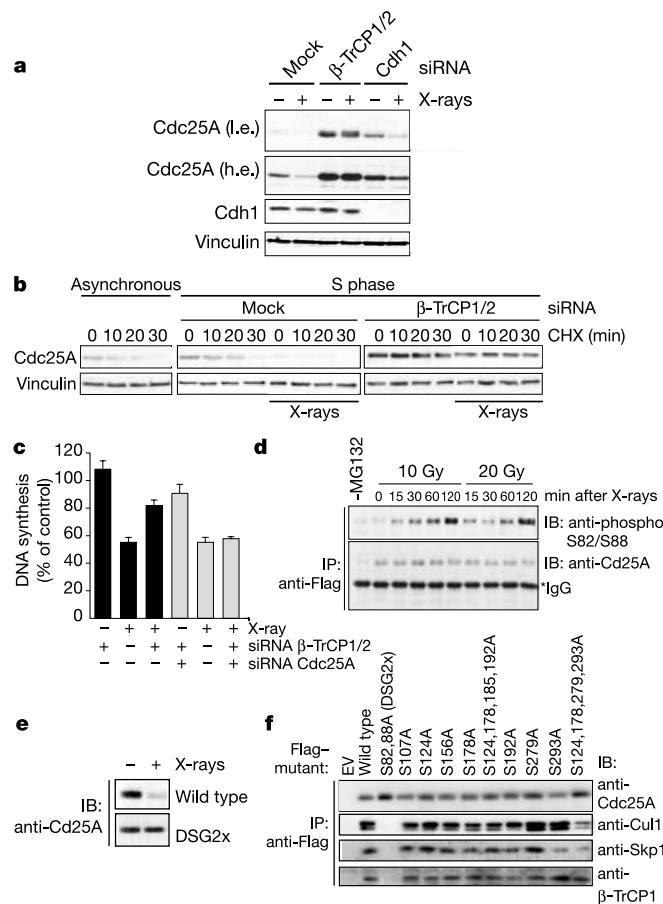


Figure 4 β -TrCP is required for degradation of Cdc25A induced by ionizing radiation in the intra-S-phase checkpoint. **a**, HeLa cells were mock-transfected or transfected with β -TrCP or Cdh1 siRNA and after 48 h were exposed to ionizing radiation (10 Gy). Cdc25A protein is shown at low and high exposure. **b**, siRNA-transfected S-phase cells were exposed to ionizing radiation, and the half-life of Cdc25A was analysed by cycloheximide (CHX) treatment. **c**, Percentage of DNA synthesis, normalized against mock-transfected non-irradiated cells, was assessed in mock-transfected cells and in cells transfected with β -TrCP1/2 or β -TrCP1/2 plus Cdc25A siRNA, 90 min after ionizing radiation treatment. **d**, HeLa cells overexpressing Flag–Cdc25A were irradiated with 10 and 20 Gy in the presence of the proteasome inhibitor MG132 and collected at the indicated time points after irradiation. Immunoprecipitated Cdc25A was immunoblotted with a purified anti-phosphoS82/S88 antibody. **e**, U2OS cells stably expressing wild-type Cdc25A or Cdc25A^{DSG2x} were mock-treated or treated by ionizing radiation. **f**, Flag-tagged Cdc25A immunocomplexes were immunoblotted for Cdc25A, Cul1, Skp1 and β -TrCP1.

Plasmids

Flag- and His-tagged Cdc25A mutants were generated using the QuickChange Site-directed Mutagenesis kit (Stratagene). We verified all constructs by DNA sequencing.

Immunoblotting, immunoprecipitation and phosphatase treatment

Immunoblotting and immunoprecipitation were done essentially as described⁸, with the following antibodies: anti-Cdc25A (F6, Santa Cruz Biotech); anti-Flag (M2, Sigma); anti-Cul1 (Zymed); anti-cyclinB1 (GNS1, Santa Cruz Biotech); anti-cyclinA (H-432, Santa Cruz Biotech); anti-Myc (9E10, Santa Cruz Biotech); anti-vinculin (Sigma); anti-Skp1 (1C10F4, Zymed); anti-β-TrCP1 (M.P.'s polyclonal serum); anti-β-TrCP2 (N-15, Santa Cruz Biotech); Anti-Emi1 antibodies were provided by P. K. Jackson (Stanford Univ. School of Medicine), and anti-Cdh1 monoclonal antibodies were supplied by K. Helin (European Inst. Oncology).

For phosphatase treatment, 500 units of λ protein phosphatase (New England Biolabs) was added to β-TrCP immunocomplexes in the presence of MgCl₂ for 30 min at 30 °C.

Peptide-binding assay

We obtained two peptides corresponding to the amino acid sequence TDSGFCLDSPGLD of human Cdc25A (Eurogentec), one of which double-phosphorylated on serine residues. The peptides were coupled to agarose beads using the Aminolink kit (Pierce). Coupled Cdc25A peptides (10 μg) were incubated with [³⁵S]methionine-labelled *in vitro* translated β-TrCP1/2 obtained by the TNT-coupled Reticulocyte Lysate system (Promega) in the presence of 5 μCi of [³⁵S]methionine (Amersham Biosciences). We washed agarose beads with RIPA buffer and assayed binding by SDS–PAGE followed by autoradiography.

In vitro ubiquitination assay

Ubiquitin ligation was determined essentially as described²⁸ by using [³⁵S]methionine-labelled *in vitro* translated Cdc25A. Baculovirus β-TrCP1, Skp2 or Fbw7 were all coexpressed with His₆-Skp1, purified by nickel-agarose chromatography and added in roughly similar amounts to the reaction.

siRNA

We purchased Cdh1 (ref. 8), β-TrCP1/2 (refs 19, 20), Emi1 (ref. 21) and Cdc25A (ref. 2) 21-base-pair siRNA oligonucleotides (Dharmacon Research). Cells were transfected with siRNA duplexes by using Oligofectamine (Invitrogen) according to the manufacturer's instructions.

Radio-resistant DNA synthesis assay

Cells were transfected with siRNA duplexes, and after 4 h were labelled for 24 h with 20 nCi ml⁻¹ [¹⁴C]thymidine (Amersham Biosciences) and then incubated in non-radioactive medium for another 24 h. Cells were irradiated with 10 Gy, incubated for 90 min at 37 °C, and pulse-labelled with 5 μCi ml⁻¹ [³H]thymidine (Amersham Biosciences) for the last 20 min. The medium was removed and the cells were washed twice with PBS, once with 5% trichloroacetic acid and twice with 100% ethanol, and then lysed with 200 μl of buffer containing 1% SDS and 10 mM NaOH. We assayed 150-μl aliquots in a liquid scintillation counter. The resulting ratios of ³H counts per min to ¹⁴C counts per min represented a measure of the DNA synthesis.

Received 17 June; accepted 17 September 2003; doi:10.1038/nature02082.

1. Falck, J., Maitland, N., Syljuasen, R. G., Bartek, J. & Lukas, J. The ATM–Chk2–Cdc25A checkpoint pathway guards against radioresistant DNA synthesis. *Nature* **410**, 842–847 (2001).
2. Zhao, H., Watkins, J. L. & Pwnica-Worms, H. Disruption of the checkpoint kinase 1/cell division cycle 25A pathway abrogates ionizing radiation-induced S and G2 checkpoints. *Proc. Natl Acad. Sci. USA* **99**, 14795–14800 (2002).
3. Sorensen, C. S. *et al.* Chk1 regulates the S phase checkpoint by coupling the physiological turnover and ionizing radiation-induced accelerated proteolysis of Cdc25A. *Cancer Cell* **3**, 247–258 (2003).
4. Maitland, N. *et al.* Rapid destruction of human Cdc25A in response to DNA damage. *Science* **288**, 1425–1429 (2000).
5. Molinari, M., Mercurio, C., Dominguez, J., Goubin, F. & Draetta, G. F. Human Cdc25 A inactivation in response to S phase inhibition and its role in preventing premature mitosis. *EMBO Rep.* **1**, 71–79 (2000).
6. Bartek, J. & Lukas, J. Mammalian G1- and S-phase checkpoints in response to DNA damage. *Curr. Opin. Cell Biol.* **13**, 738–747 (2001).
7. Falck, J., Petrini, J. H., Williams, B. R., Lukas, J. & Bartek, J. The DNA damage-dependent intra-S phase checkpoint is regulated by parallel pathways. *Nature Genet.* **30**, 290–294 (2002).
8. Donzelli, M. *et al.* Dual mode of degradation of Cdc25A phosphatase. *EMBO J.* **21**, 4875–4884 (2002).
9. Jackson, P. K. & Eldridge, A. G. The SCF ubiquitin ligase: an extended look. *Mol. Cell* **9**, 923–925 (2002).
10. Patton, E. E., Willems, A. R. & Tyers, M. Combinatorial control in ubiquitin-dependent proteolysis: don't skip the F-box hypothesis. *Trends Genet.* **14**, 236–243 (1998).
11. Spruck, C. H. & Strohmaier, H. M. Seek and destroy: SCF ubiquitin ligases in mammalian cell cycle control. *Cell Cycle* **1**, 250–254 (2002).
12. Cenciarelli, C. *et al.* Identification of a family of human F-box proteins. *Curr. Biol.* **9**, 1177–1179 (1999).
13. Kipreos, E. T. & Pagano, M. The F-box protein family. *Genome Biol.* **1**, Reviews3002.1–3002.7 [online] (<http://genomebiology.com/2000/1/5/reviews/3002>) (2000).
14. Yaron, A. *et al.* Identification of the receptor component of the IκBα-ubiquitin ligase. *Nature* **396**, 590–594 (1998).
15. Winston, J. T. *et al.* The SCF β-TrCP-ubiquitin ligase complex associates specifically with phosphorylated destruction motifs in IκBα and β-catenin and stimulates IκBα ubiquitination *in vitro*. *Genes Dev.* **13**, 270–283 (1999).

16. Lassot, I. *et al.* ATF4 degradation relies on a phosphorylation-dependent interaction with the SCF^{β-TrCP} ubiquitin ligase. *Mol. Cell. Biol.* **21**, 2192–2202 (2001).
17. Lang, V. *et al.* β-TrCP-mediated proteolysis of NF-κB1 p105 requires phosphorylation of p105 serines 927 and 932. *Mol. Cell. Biol.* **23**, 402–413 (2003).
18. Liu, C. *et al.* Control of β-catenin phosphorylation/degradation by a dual-kinase mechanism. *Cell* **108**, 837–847 (2002).
19. Guardavaccaro, D. *et al.* Control of meiotic and mitotic progression by the F box protein β-Trcp1 *in vivo*. *Dev. Cell* **4**, 799–812 (2003).
20. Margottin-Goguet, F. *et al.* Prophase destruction of Emi1 by the SCF^{β-TrCP/Slimb} ubiquitin ligase activates the anaphase promoting complex to allow progression beyond prometaphase. *Dev. Cell* **4**, 813–826 (2003).
21. Hsu, J. Y., Reimann, J. D., Sorensen, C. S., Lukas, J. & Jackson, P. K. E2F-dependent accumulation of hEmi1 regulates S phase entry by inhibiting APC^{Cdh1}. *Nature Cell Biol.* **4**, 358–366 (2002).
22. Painter, R. B. & Young, B. R. Radiosensitivity in ataxia-telangiectasia: a new explanation. *Proc. Natl Acad. Sci. USA* **77**, 7315–7317 (1980).
23. Verma, R. *et al.* Phosphorylation of Sic1p by G1 Cdk required for its degradation and entry into S phase. *Science* **278**, 455–460 (1997).
24. Nash, P. *et al.* Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* **414**, 514–521 (2001).
25. Wu, G. *et al.* Structure of a β-TrCP1–Skp1–β-catenin complex: destruction motif binding and lysine specificity of the SCF^{β-TrCP1} ubiquitin ligase. *Mol. Cell* **11**, 1445–1456 (2003).
26. Clucas, C., Cabello, J., Bussing, I., Schnabel, R. & Johnstone, I. L. Oncogenic potential of a *C. elegans* *cdc25* gene is demonstrated by a gain-of-function allele. *EMBO J.* **21**, 665–674 (2002).
27. Bell, D. W. *et al.* Heterozygous germ line hCHK2 mutations in Li–Fraumeni syndrome. *Science* **286**, 2528–2531 (1999).
28. Carrano, A. C., Eytan, E., Hershko, A. & Pagano, M. SKP2 is required for ubiquitin-mediated degradation of the CDK inhibitor p27. *Nature Cell Biol.* **1**, 193–199 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank T. K. Ko for his contribution; Z. G. Pan and S. Reed for reagents; G. Ossolengo for producing and purifying the anti-phosphoS82/S88 antibody; and M. Squatrito and E. De Billy for discussion. This work was supported by an Italian American Cancer Foundation fellowship and a Susan Komen Breast Cancer Foundation fellowship to D.G.; a Irma Hirsch Scholarship and grants from the NIH to M.P.; and grants from the Italian Association for Cancer Research (AIRC), the Italian Foundation for Cancer Research (FIRC) and Telethon to G.F.D.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.D. (mdonzell@ieo.it).

Structural adaptability in the ligand-binding pocket of the ecdysone hormone receptor

Isabelle M. L. Billas, Thomas Iwema, Jean-Marie Garnier, André Mitschler, Natacha Rochel & Dino Moras

Département de Biologie et de Génomique Structurales, IGBMC, CNRS/INSERM/Université Louis Pasteur, Parc d'Innovation BP10142, 67404 Illkirch cedex, France

The ecdysteroid hormones coordinate the major stages of insect development, notably moulting and metamorphosis, by binding to the ecdysone receptor (EcR); a ligand-inducible nuclear transcription factor^{1,2}. To bind either ligand or DNA, EcR must form a heterodimer with ultraspiracle (USP), the homologue of retinoid-X receptor^{3–5}. Here we report the crystal structures of the ligand-binding domains of the moth *Heliothis virescens* EcR–USP heterodimer in complex with the ecdysteroid ponasterone A and with a non-steroidal, lepidopteran-specific agonist BY106830 used in agrochemical pest control. The two structures of EcR–USP emphasize the universality of heterodimerization as a general mechanism common to both vertebrates and invertebrates. Comparison of the EcR structures in complex with steroidal and non-steroidal ligands reveals radically different and only partially overlapping ligand-binding pockets that

could not be predicted by molecular modelling and docking studies^{6,7}. These findings offer new perspectives for the design of insect-specific, environmentally safe insecticides. The concept of a ligand-dependent binding pocket in EcR provides an insight into the moulding of nuclear receptors to their ligand, and has potential applications for human nuclear receptors.

The EcR has been an important target in the design of new, environmentally safe insecticides against pest species that cause billions of pounds of damage to global agriculture each year. The bisacylhydrazines⁸, a class of synthetic agonists structurally and chemically different from ecdysteroids, were reported to induce premature and lethal moulting in various types of caterpillar pest larvae. For our structural study of the *H. virescens* (Lepidoptera) EcR–USP heterodimer, we used the ecdysteroid ponasterone A (ponA) as well as the non-steroidal bisacylhydrazine agonist BYI06830 (Supplementary Fig. 4b), a lepidopteran-specific compound similar to chromafenozide⁹. Crystals were obtained for the ligand-binding domains (LBDs) of the *H. virescens* EcR–USP heterodimer (HvEcR-LBD–HvUSP-LBD) in complex with ponA, and for the more soluble mutant HvEcR-LBD(W303Y/A361S/L456S/C483S)–HvUSP-LBD in complex with BYI06830. The mutated residues are located at the surface of the LBD and do not affect the structure of the receptor, its ability to heterodimerize, or its function, as demonstrated by transactivation experiments and

mass spectrometric measurements (Supplementary Fig. 4).

The overall structures of both EcR–USP heterodimers (that is, when bound to ponA or to BYI06830) are comparable (Fig. 1a, b). Their butterfly shape is similar to that seen for the vertebrate heterodimer complexes PPAR γ –RXR α , RAR α –RXR α and LXR α –RXR β ^{10–12}. The general architecture of the EcR and USP LBDs composing each heterodimer is similar to that seen in the crystal structures of other nuclear receptor LBDs, with a general fold consisting of a three-layered, antiparallel, α -helical sandwich and a β -sheet. The structures of both HvUSP-LBDs are virtually identical to those determined previously for the isolated HvUSP-LBD^{13,14}. For HvUSP-LBD bound to either the steroidal or non-steroidal ligand the H12 helix, which bears the ligand-dependent activation function (AF-2), is in a so-called antagonist conformation. In contrast, H12 of both agonist-bound EcR-LBDs adopts the canonical active conformation. These two heterodimeric complexes are thus the first crystal structures of nuclear receptor heterodimers with a dissymmetry in the H12 conformations^{10–12}. The EcR–USP heterodimer interface is similar to that of vertebrate heterodimers and involves H7, H9, H10 and the loop between H8 and H9 of both LBDs^{10–12}. The novel feature concerns the long peptide inserted between H5 and the β -sheet in HvUSP, which is partially stabilized. The electron density suggests possible contacts between this insertion and the loop H8–H9 of the EcR-LBD partner.

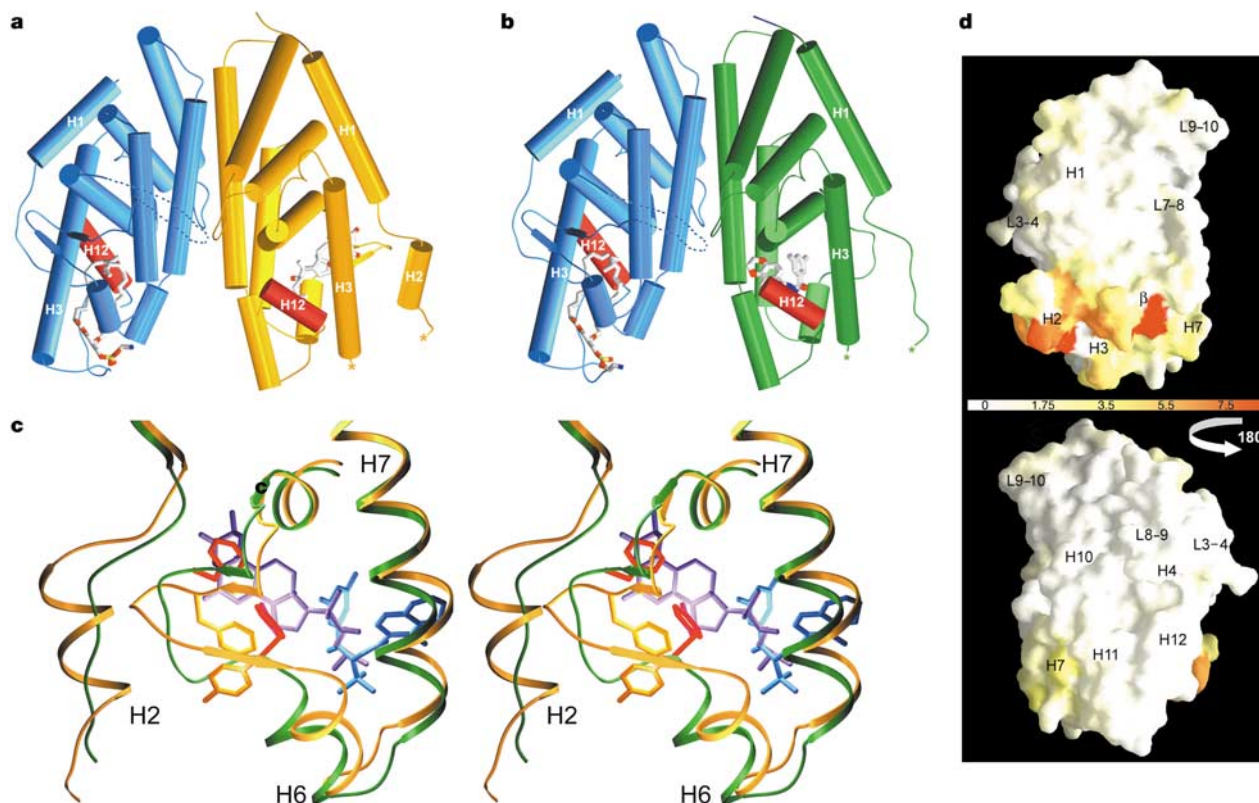


Figure 1 The two structures of the ligand-binding domains of EcR–USP. **a, b**, Overall structure of the EcR–USP heterodimers in complex with the steroid hormone ponA (**a**) and the non-steroidal ligand BYI06830 (**b**). Helices H12 are shown as red cylinders. For USP, the missing loop connecting H5 to the β -sheet is shown by a blue dashed line going close to the EcR H9–H10 loop, as suggested by the discontinuous electron density. The EcR ligands and the phospholipid in USP are shown in stick representation with carbon coloured in white, oxygen in red and nitrogen in blue. **c**, Ribbon diagram showing the superimposition of the ponA- and BYI06830-bound EcR-LBD structures. The view is restricted to the region differing the most between the two EcR-LBD structures

encompassing H2, H6, H7 and the β -sheet. The ponA-bound EcR-LBD is shown in yellow with ponA in magenta and the BYI06830-bound EcR-LBD in green with BYI06830 in blue. The two residues discussed in the text, F397 and Y403, are shown in yellow for the ponA-bound EcR-LBD and in red for the BYI06830-bound EcR-LBD. **d**, Surface representations of the ponA-bound EcR-LBD coloured according to the r.m.s. deviations calculated residue by residue between the ponA- and the BYI06830-bound EcR-LBDs. The scale ranges from 0 (white) to 7.5 Å (red). The two views are related by a 180° rotation around the vertical axis.

As the H8–H9 loop represents an essential element of the heterodimer interface, the USP-specific H5- β -sheet insertion might have an additional yet unknown regulatory role.

The EcR-LBD structures are most closely related to the $1\alpha,25$ -dihydroxyvitamin D₃ receptor (VDR)-LBD structure, with a root-mean-square (r.m.s.) deviation of 1.2 Å over 197 C α positions for the ponA-bound EcR-LBD and 1.1 Å over 182 C α positions for the BYI06830-bound EcR-LBD. The superimposition of the two HvEcR-LBD structures reveals marked differences, mostly affecting the region comprising H6, H7, the β -sheet and the loop between H1 and H3, where an additional helix H2 is observed for the ecdysteroid-bound EcR-LBD. Substantial shifts of H6 and the amino-terminal part of H7 are observed, whereas the β -sheet is markedly affected owing to the disruption of interactions between the second and third strands (Fig. 1c, d). It is interesting to note the proximity of this region to the carboxy terminus of H1, which is implicated in cofactor binding¹⁵.

Along with the differences observed in the receptor scaffold, the EcR-LBDs in complex with the steroidal and non-steroidal agonists exhibit different and only partially overlapping ligand-binding cavities (Fig. 2a). For EcR-LBD in complex with ponA, the long and thin L-shaped cavity is quite similar to the VDR-LBD ligand-binding pocket¹⁶. This correspondence is further emphasized by the superimposition of ponA and $1\alpha,25$ -dihydroxyvitamin D₃ bound to their cognate receptors, revealing a similar positioning of their respective aliphatic side chain (at position 17 of the D-ring) (Fig. 2b). In contrast with the ponA-bound EcR cavity, which is completely buried inside the receptor LBD, the ligand-binding pocket of the BYI06820-bound EcR-LBD consists of a rather bulky V-shaped cavity with an open cleft between H7 and H10. This opening extends towards the H8–H9 loop of the heterodimer partner USP. The superposition of the steroidal and non-steroidal ligands as found in the EcR cavities indicates that the *t*-butyl group together with the benzoyl A-ring of BYI06830 superimpose to the hydroxylated side chain of ponA at the C17 position (Fig. 2c). This superimposition severely contrasts with predictions of modelling studies that considered a single binding niche^{6,7}.

The key hormone for the development of most insects is 20-hydroxyecdysone (20E), but experimental evidence indicates that ecdysone and other ecdysteroids coexist with 20E and have distinct

roles at different insect developmental stages^{17,18}. The ecdysteroid ponA is identical to 20E but lacks the 25-OH group. It was chosen in our crystallization experiments for its higher affinity compared with 20E. Therefore, the crystal structure of ponA-bound EcR-LBD allows one to gain insight into the more general binding mode of ecdysteroids to EcR (Fig. 3a, c). In particular, the strong hydrogen bond observed between the 20-hydroxyl group and Y408 provides a rationale for the higher activity of 20E compared with ecdysone as ascertained through *in vitro* studies¹⁹. Furthermore, examination of the ligand-binding pocket shows that most residues interacting with ponA are conserved among species, rationalizing the promiscuous character of 20E (Supplementary Fig. 5). PonA is directly involved in the stabilization of the EcR-LBD structure. In fact, the helical conformation of H2 is stabilized by interactions between the ponA C2- and C3-hydroxyl groups and residues of the H1–H2 loop, helix H5 and the β -sheet.

For EcR-LBD in complex with BYI06830, the loss of scaffolding interactions that exist in the ponA-bound EcR-LBD destabilizes the helical conformation of H2. Furthermore, binding of the synthetic ligand leads to disruption of the interactions between the second and third β -sheet strands. The structural reorganization of the β -sheet is mainly due to the outer to inner motion of F397 and Y403, located on each side of the β -sheet tip. These residues fill the region of the cavity occupied by the A-, B- and C-rings of the ecdysteroid; however, the region is left empty by the ligand when EcR-LBD is bound to BYI06830 (Figs 1c and 3b, d). Three polar amino acid residues (T343, N504, Y408) are implicated in the hydrogen-bond interaction network with the ligand carbonyl groups and the unsubstituted amide nitrogen. On the other hand, the *t*-butyl is located in a hydrophobic part of the cavity formed by residues belonging to H3, H11 and to the loops H6–H7 and H11–H12. The presence of this group in the dibenzoylhydrazine class of compounds confers a high ecdysone-like activity²⁰. Examination of the cavity residues allows us to rationalize the specificity of BYI06830 to Lepidoptera. In particular, residue V384 in H5—conserved in lepidopteran insects but replaced by methionine in other insects—is essential for this species specificity. In fact, all possible methionine rotamers would sterically interfere with either the BYI06830 methyl groups on the A-ring or with the protein itself. Therefore, this cavity represents a good template for the design of

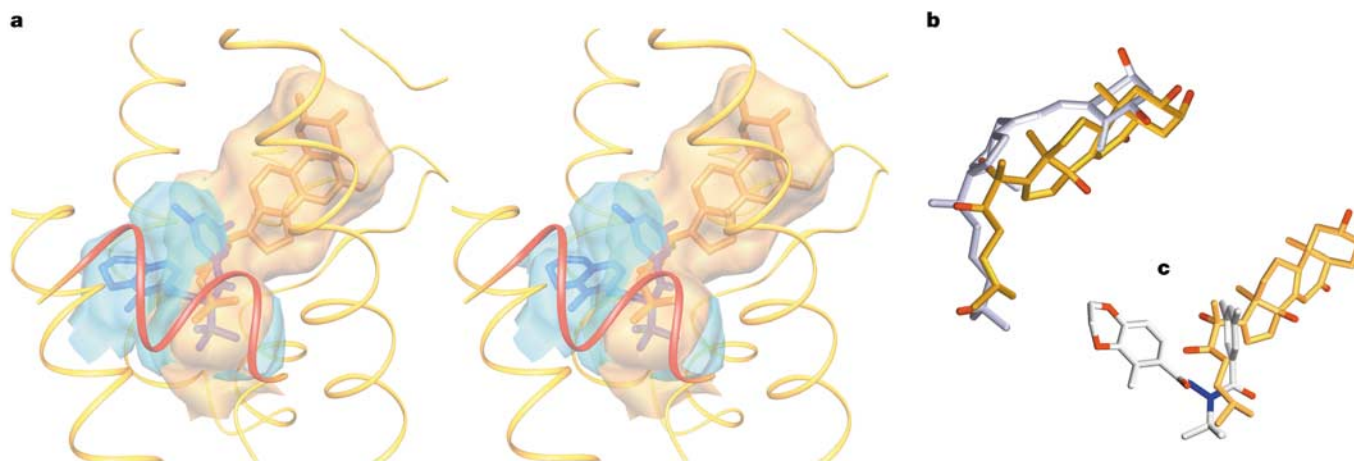


Figure 2 The LBDs of EcR complexed to a steroidal and a non-steroidal ligand exhibit different and only partially overlapping ligand-binding cavities. **a**, Stereoview of the two ligand-binding cavities of EcR-LBD, together with their respective ligand. The view is identical to that of Fig. 1a, b. The ponA-bound EcR cavity is shown in light orange and the BYI06830-bound EcR cavity in blue. The ligands are shown as stick models with ponA in orange and BYI06830 in blue. **b**, Superimposition of ponA and $1\alpha,25$ -dihydroxyvitamin

D₃ bound to their cognate receptor. The ponA-bound EcR-LBD and the vitamin D-bound VDR-LBD were superimposed using Lsq-man²⁸. **c**, Superimposition of the steroidal and non-steroidal EcR ligands as bound to the EcR-LBD. Atom colouring is red for oxygen, blue for nitrogen, yellow for carbon in ponA and grey for carbon in vitamin D and in BYI06830.

species-specific synthetic agonists. However, in the case of synthetic ligands that are different from bisacylhydrazines, the adaptability of EcR to its ligand might give rise to yet another ligand-binding pocket.

To assess further the functional relevance of the ligand-protein contacts, we individually mutated key interacting residues to alanine (Fig. 3e). Our results are consistent with predictions based on the EcR-LBD structures. For example, the marked reduction of transcriptional activity for the N504A and W526A mutants can be explained by their structural role in ligand binding and ligand-

mediated stabilization of the agonist H12 conformation. These results agree with recent observations of the dominant-negative phenotype of the *Drosophila melanogaster* EcR(W650A) mutant (corresponding to HvEcR(W526A)), for which neither ligand-binding nor transcriptional activation are observed²¹. Furthermore, our data are consistent with recently published functional studies of the lepidopteran *Choristoneura fumiferana* (Cf)EcR single mutants, in particular with the ecdysteroid-selective mutants R378A, F392A and A393P (R383A, F397A and A398P for HvEcR). The CfEcR(A393P) mutant is particularly interesting for its critical role

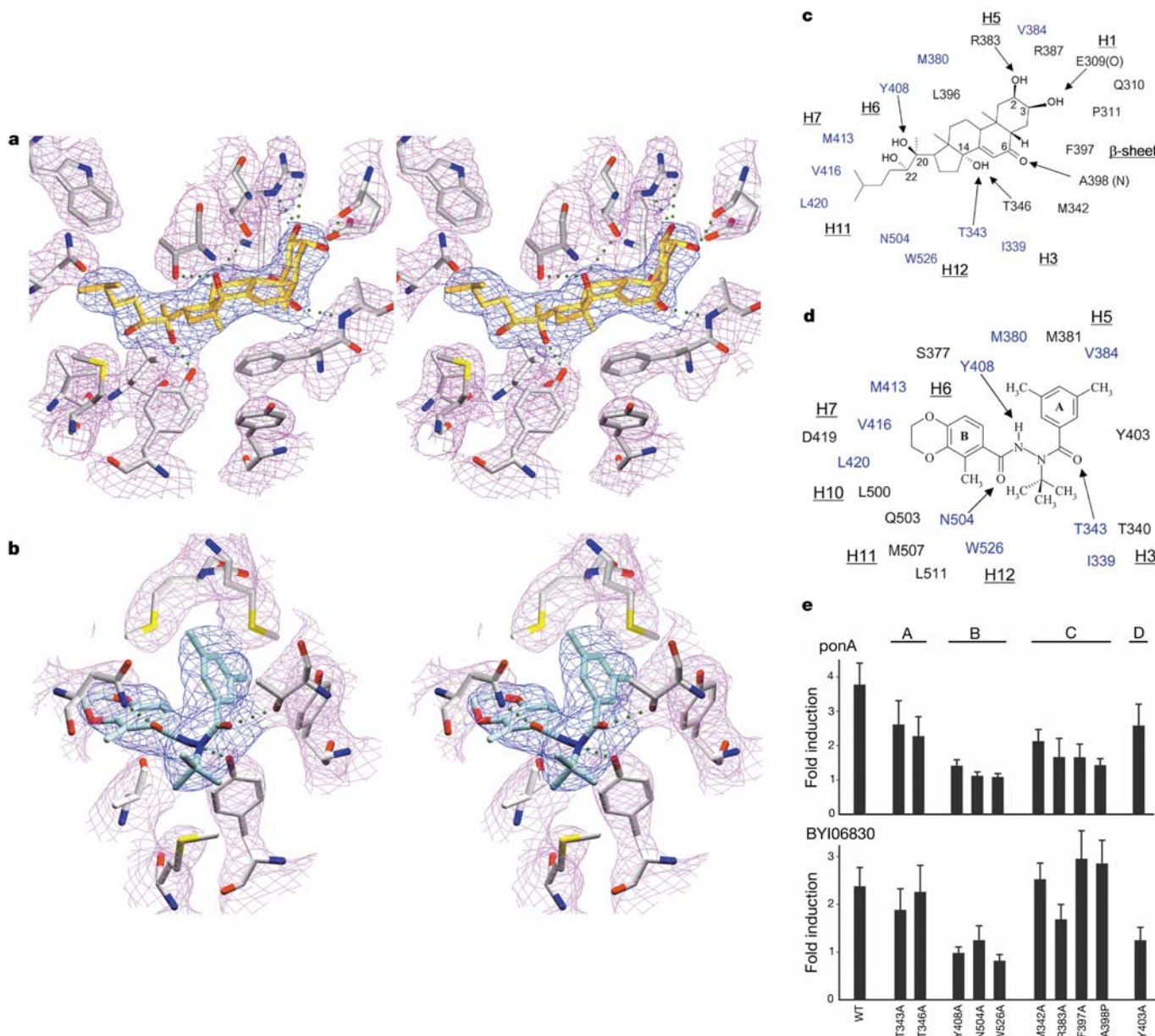


Figure 3 Two radically different binding modes for EcR. **a, b**, Stereoviews of the electron density for the ponA-bound EcR-LBD at 2.9 Å resolution (**a**) and the BY106830-bound EcR-LBD at 3.0 Å resolution (**b**). Two different maps are shown in each figure: a sigma-A-weighted $2F_o - F_c$ omit map for the ligand (in blue) and a sigma-A-weighted $2F_o - F_c$ map (in magenta) for selected residues in the ligand-binding pocket. In both cases the map is contoured at 1σ and is overlaid on the final refined models. Hydrogen bonds between ligand and residues are indicated by green dotted lines. **c, d**, Schematic representation of the interactions of ponA (**c**) and BY106830 (**d**) with the binding cavity.

Arrows correspond to hydrogen bonds between ligand and amino acid residues. Residues in blue are common to both structures. **e**, Differential effects of mutated residues in the binding cavities of HvEcR. Mutants T343A and T346 (A) show a small reduction of the transcriptional activity for both ligands; Y408A, N504A and W526A (B) are null mutants for both ligands; M342A, R383A, F397A and A398P (C) show specific reduction in ecdysteroid transcriptional activity; and Y403A (D) affects more specifically the transcriptional activity induced by BY106830.

in discriminating between steroidal and non-steroidal ligands²². For the ponA-bound HvEcR-LBD, A398 is located at the tip of the β -sheet and is in close contact with the ligand, whereas for BY106830-bound HvEcR-LBD, the β -sheet region including A398 is not involved in ligand recognition. Mutation of A398 to a constraining proline residue affects the β -sheet structure by introducing local distortions of the protein backbone that can explain the complete loss of activity for the steroidal, but not for the non-steroidal, ligand.

The ability of EcR-LBD to adopt different binding cavities highlights the extreme flexibility and adaptability of this protein domain, properties that allow the moulding of the receptor around its ligand and consequently the formation of the proper ligand-binding pocket. These findings introduce a new concept for the field of nuclear receptors; that is, the ligand-dependent binding pocket. Indeed, up to now the structures of other nuclear receptor LBDs have indicated that small local adaptations of the receptor LBD to its ligand can take place. These modifications might lead to a substantial variation in the size of the pocket, but do not affect the structure of the LBD, as shown recently for the liver X receptor²³. However, unlike in our case, these cavity modifications never lead to a ligand-dependent pocket with different residues implicated in ligand recognition. Our observations further rationalize the need of either chaperones²⁴ or USP for the stabilization of EcR in solution. In fact, EcR needs to heterodimerize with its partner USP for the stabilization of a ligand-binding configuration. Our structural work suggests that ligand binding requires a portion of the EcR-LBD to be stiffened by the anchoring of USP through interface contacts, allowing the more flexible region of the LBD to mould around the ligand. The outer to inner switch of F397 and Y403, which participate in the adaptation mechanism (Fig. 1c), strongly suggests that entropic effects and correlated desolvation processes are essential for the structural adaptation of the receptor to its ligand; a phenomenon already observed in the structural transition from the apo- to the holo-forms of RXR²⁵.

The crystal structures of EcR-LBD in complex with a steroidal and a non-steroidal ligand raise several points of general interest for the superfamily of nuclear receptors. First, the structural differences in the receptor LBD provide an indirect experimental observation of the interaction region between co-activators and receptors. Both structures indicate that co-activator binding to the LXXLL hydrophobic groove of the receptor is unaltered, consistent with the fact that both ligands are good agonists. However, the difference in biological action of synthetic agonists compared with ecdysteroids^{8,9} suggests that some cofactors might be at least partially discriminated owing to a different receptor conformation in the region between H1, H3 and the β -sheet. Most notably, our findings provide experimental evidence that nuclear receptors and their ligands can mutually adapt and fit to each other. Other examples of this type of structural adaptability probably exist and should offer new perspectives for drug design. Candidates for this type of behaviour might be found among steroid receptors such as the mineralocorticoid and glucocorticoid receptors, which are known to be dependent on chaperones for stabilization. Furthermore, our data give strong support to unbiased experimental ligand search for both ligand-regulated and constitutively active orphan nuclear receptors. The search for new types of therapeutic agonists might therefore benefit from these unforeseen prospects. □

Methods

Protein preparation

HvUSP-LBD (residues 205–466) was expressed from the T7 promoter of plasmid vector pACYC11b. Wild-type HvEcR-LBD (residues 284–532) was expressed as a His₆-thioredoxine fusion protein from the expression vector pET32b. Mutant HvEcR-LBD(W303Y/A361S/L456S/C483S) (residues 285–532) was expressed either as an N-terminal His₆-tagged fusion protein (pET28b) or as a His₆-thioredoxine fusion protein (pET32b). BL21(DE3) cells transformed with the USP-LBD and EcR-LBD expression plasmids were grown in two-times LB medium at 37 °C in the presence of 30 μ M ligand

and subsequently induced for 18 h with 0.8 mM isopropyl- β -D-thiogalactopyranoside at 15 °C. Cell pellets were resuspended, lysed by sonication in the presence of 30 μ M ligand, and clarified by centrifugation. The cleared supernatant was loaded onto a cobalt-chelating resin (TALON Superflow, Clontech). The heterodimer complex was eluted using 250 mM imidazole (pH 8.0). After tag removal by thrombin digestion the protein was subjected to gel filtration on a Superdex 200 16/60 column (Pharmacia Biotech). The 1:1 molar ratio of EcR-LBD and USP-LBD was confirmed by N-terminal amino-acid sequencing, SDS-polyacrylamide gel electrophoresis analysis and electrospray ionization mass spectrometry.

Crystallization

The purified ponA- and BY106830-bound heterodimer complexes were conditioned in 25 mM Tris (pH 8.0), 100 mM NaCl, 100 mM KCl, 4 mM CHAPS, 2% glycerol (v/v), 500 mM ligand and concentrated to 4–6 mg ml⁻¹. Crystallization experiments were carried out at 24 °C using both the hanging-drop vapour and the sitting-drop diffusion methods. The crystallization conditions were 30% PEG 4000, 0.2 M ammonium sulphate and 0.1 M Na citrate (pH 5.6) for ponA-bound HvEcR-LBD–HvUSP-LBD, and 10% PEG 1000, 10% PEG 8000, 0.3 M MgCl₂ and 0.1 M HEPES (pH 7.5) for BY106830-bound HvEcR-LBD(W303Y/A361S/L456S/C483S)–HvUSP-LBD. Crystals were flash-frozen in liquid nitrogen after a short dip in a solution containing 15% ethylene glycol and 5% PEG 400 as cryoprotectant.

Data collection, structure determination and refinement

The crystal structures were solved by molecular replacement with EPMR²⁶ using the crystal structure of VDR-LBD¹⁶ (Protein Data Bank code 1DB1) as a search model (26% sequence identity). Clear solutions were obtained in the space group *P*₃2₁ for the ponA complex and *P*321 for the BY106830 complex. In both cases the asymmetric unit contained one EcR-LBD–USP-LBD heterodimer. Refinement was performed with CNS using torsion angle dynamics with a maximum likelihood function target, including restrained anisotropic refinement of individual atomic temperature factors and bulk solvent correction²⁷. Manual adjustments and rebuilding of the models were performed using the program O²⁸ and sigma-A-weighted electron density maps. The final model for ponA-bound EcR-LBD–USP-LBD comprises residues V287 to G323 and D331 to A529 for EcR and V205 to A304 and A315 to R455 for USP. The final model for BY106830-bound EcR-LBD–USP-LBD comprises residues V287 to R318 and D333 to A528 for EcR and A206 to A304 and G316 to R455 for USP. The missing regions correspond to flexible loops between H1 and H3 in EcR and between H5 and the β -sheet for USP and to the chain termini. Residues Q321, T322, W323, R337, K517 and V528 (EcR) and E304 and R315 (USP) for the ponA complex, and residues D389, R482, K517 and V528 (EcR) and Q206, E207, E304, T316, T317 and S318 (USP) for the BY106830 complex were poorly resolved in the electron-density maps, and were modelled as alanines or glycines. Structures exhibit good geometry with no Ramachandran outliers. Molecular graphics figures were created with SETOR, DINO and GRASP programs.

Chemicals

Ponasterone A was purchased from SciTech and BY106830 from Bayer CropScience, and they were dissolved in ethanol and dimethylsulphoxide, respectively.

Cell culture

Human embryonic kidney 293 EBNA cells were cultured in Dulbecco's modified Eagle's medium. The medium was supplemented with 10% heat-inactivated fetal calf serum, gentamycin and G418 sulphate at 37 °C under 5% CO₂ humidified atmosphere.

Transfection assays

Transient transfection assays were carried out in 24-well plates (10⁵ cells per well) using a standard calcium phosphate co-precipitation technique as described previously²⁹. Cells were co-transfected with 25 ng pSG5-HvEcR wild-type or pSG5-HvEcR mutant expression plasmid and 25 ng pSG5-HvUSP expression plasmid, 250 ng PAL1(x5)-TK-LUC reporter and 75 ng pCH110 β -galactosidase expression plasmid. Cells were treated with either ponA or BY106830 for 20 h, and cell extracts were assayed for luciferase and β -galactosidase activity. Luciferase values were normalized to β -galactosidase activity. Data points represent the mean of assays performed in duplicate for at least three independent experiments. Mutants were generated using the QuickChange site-directed mutagenesis kit from Stratagene. All mutants were confirmed by sequence analysis.

Electrospray ionization mass spectrometry

The protein HvEcR-LBD–HvUSP-LBD was concentrated to about 3 mg ml⁻¹ and dialysed against 200 mM NH₄OAc at pH 6.5 using centricon-50 concentrators (Amicon), and diluted to a final concentration of 10⁻⁵ M. Studies were performed using an electrospray ionization time-of-flight mass spectrometer (LCT, Micromass) and continuously infusing the sample into the ion source at a flow rate of 4 μ l min⁻¹ with a Harvard Model 11 syringe pump (Harvard Apparatus). The detection of ligand–receptor non-covalent complexes was achieved according to the procedure described in ref. 30.

Received 24 July; accepted 14 October 2003; doi:10.1038/nature02112.

Published online 2 November 2003.

1. Koelle, M. R. *et al.* The *Drosophila* EcR gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily. *Cell* **67**, 59–77 (1991).
2. Riddiford, L. M., Cherbas, P. & Truman, J. W. Ecdysone receptors and their biological actions. *Vitam. Horm.* **60**, 1–73 (2001).

3. Yao, T.-P., Segreaves, W. A., Oro, A. E., McKeown, M. & Evans, R. M. *Drosophila* ultraspiracle modulates ecdysone receptor function via heterodimer formation. *Cell* **71**, 63–72 (1992).
4. Thomas, H. E., Stunnenberg, H. G. & Stewart, A. F. Heterodimerization of the *Drosophila* ecdysone receptor with retinoid X receptor and ultraspiracle. *Nature* **362**, 471–475 (1993).
5. Yao, T.-P. *et al.* Functional ecdysone receptor is the product of EcR and ultraspiracle genes. *Nature* **366**, 476–479 (1993).
6. Wurtz, J.-M. *et al.* A new model for 20-hydroxyecdysone and dibenzoylhydrazine binding: A homology modeling and docking approach. *Protein Sci.* **9**, 1073–1084 (2000).
7. Kasuya, K. *et al.* Binding mode of ecdysone agonists to the receptor: comparative modeling and docking studies. *J. Mol. Model.* **9**, 58–65 (2003).
8. Dhadialla, T. S., Carlson, G. R. & Le, D. P. New insecticides with ecdysteroidal and juvenile hormone activity. *Annu. Rev. Entomol.* **43**, 545–569 (1998).
9. Toya, T., Fukasawa, H., Masui, A. & Endo, Y. Potent and selective partial ecdysone agonist activity of chromafenozide in Sf9 cells. *Biochem. Biophys. Res. Commun.* **292**, 1087–1091 (2002).
10. Gampe, R. T. Jr *et al.* Asymmetry in the PPAR γ /RXR α crystal structure reveals the molecular basis of heterodimerization among nuclear receptors. *Mol. Cell* **5**, 545–555 (2000).
11. Bourguet, W. *et al.* Crystal structure of a heterodimeric complex of RAR and RXR ligand-binding domains. *Mol. Cell* **5**, 289–298 (2000).
12. Svensson, S. *et al.* Crystal structure of the heterodimeric complex of LXR α and RXR β ligand-binding domains in a fully agonistic conformation. *EMBO J.* **22**, 4625–4633 (2003).
13. Billas, I. M. L., Moulinier, L., Rochel, N. & Moras, D. Crystal structure of the ligand binding domain of the ultraspiracle protein USP, the ortholog of RXRs in insects. *J. Biol. Chem.* **276**, 7465–7474 (2001).
14. Clayton, G. M., Peak-Chew, S. Y., Evans, R. M. & Schwabe, J. W. The structure of the ultraspiracle ligand-binding domain reveals a nuclear receptor locked in an inactive conformation. *Proc. Natl Acad. Sci. USA* **98**, 1549–1554 (2001).
15. Horlein, A. J. *et al.* Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature* **377**, 397–404 (1995).
16. Rochel, N., Wurtz, J.-M., Mitschler, A., Klaholz, B. P. & Moras, D. The crystal structure of the nuclear receptor of vitamin D bound to its natural ligand. *Mol. Cell* **5**, 173–179 (2000).
17. Henrich, V. C. & Brown, N. E. Insect nuclear receptors: A developmental and comparative perspective. *Insect Biochem. Mol. Biol.* **25**, 881–897 (1995).
18. Gilbert, L. I., Rybczynski, R. & Warren, J. T. Control and biochemical nature of the ecdysteroidogenic pathway. *Annu. Rev. Entomol.* **47**, 883–916 (2002).
19. Baker, K. D., Warren, J. T., Thummel, C. S., Gilbert, L. I. & Mangelsdorf, D. J. Transcriptional activation of the *Drosophila* ecdysone receptor by insect and plant ecdysteroids. *Insect Biochem. Mol. Biol.* **30**, 1037–1043 (2000).
20. Mohammed-Ali, A. K., Chan, T.-H., Thomas, A. W., Strunz, G. M. & Jewett, B. Structure-activity relationship study of synthetic hydrazines as ecdysone agonists in the control of spruce budworm (*Choristoneura fumiferana*). *Can. J. Chem.* **73**, 550–557 (1995).
21. Hu, X., Cherbas, L. & Cherbas, P. Transcription activation by the ecdysone receptor (EcR/USP): Identification of activation functions. *Mol. Endocrinol.* **17**, 716–731 (2003).
22. Kumar, M. B., Fujimoto, T., Potter, D. W., Deng, C. & Palli, S. R. A single point mutation in ecdysone receptor leads to increased ligand specificity: Implications for gene switch applications. *Proc. Natl Acad. Sci. USA* **99**, 14710–14715 (2002).
23. Farnegardh, M. *et al.* The three dimensional structure of the liver X receptor beta reveals a flexible ligand binding pocket that can accommodate fundamentally different ligands. *J. Biol. Chem.* **40**, 38821–38828 (2003).
24. Arbeitman, M. N. & Hogness, D. S. Molecular chaperones activate the *Drosophila* ecdysone receptor, an RXR heterodimer. *Cell* **101**, 67–77 (2000).
25. Egea, P. F. *et al.* Crystal structure of the human RXR α ligand-binding domain bound to its natural ligand 9-cis retinoic acid. *EMBO J.* **19**, 2592–2601 (2000).
26. Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr. D* **55**, 484–491 (1999).
27. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
28. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
29. Greiner, E. F. *et al.* Functional analysis of retinoid Z receptor beta, a brain-specific nuclear orphan receptor. *Proc. Natl Acad. Sci. USA* **93**, 10105–10110 (1996).
30. Potier, N. *et al.* Using non denaturing mass spectrometry to detect fortuitous ligands in orphan nuclear receptors. *Protein Sci.* **12**, 725–733 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to G. Holmwood, K. Tietjen and M. Schindler (Bayer CropScience) for providing us with BY106830 and ponA. We thank H. Greschik, H. Gronemeyer, B. Klaholz and G. Richards for critical comments on the manuscript; Y. Brelivet and J. M. Wurtz for discussion and advice; H. Nierengarten for mass spectrometry analysis; V. Chavant, D. Rose and F. Zink for technical assistance; S. Sasorith for her help with figures; and C. Massobrio for support and encouragement. We thank the staff of the ESRF ID14 beamline (Grenoble, France) and of the French beamline BM30A (Grenoble, France) for assistance during synchrotron data collection. This work was supported by Bayer CropScience and by grants from CNRS, INSERM, the Hôpital Universitaire de Strasbourg, the Ministère de la Recherche et de la Technologie (Programme de Génomique Structurale) and EU-SPINE.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.M. (moras@igbmc.u-strasbg.fr). Atomic coordinates have been deposited in the Protein Data Bank under accession codes 1R1K and 1R2O.

Crystal structure of a zinc-finger–RNA complex reveals two modes of molecular recognition

Duo Lu, M. Alexandra Searles & Aaron Klug

Medical Research Council (MRC) Laboratory of Molecular Biology, Cambridge CB2 2QH, UK

Zinc-finger proteins of the classical Cys₂His₂ type are the most frequently used class of transcription factor and account for about 3% of genes in the human genome^{1,2}. The zinc-finger motif was discovered³ during biochemical studies on the transcription factor TFIIIA, which regulates the 5S ribosomal RNA genes of *Xenopus laevis*^{4,5}. Zinc-fingers mostly interact with DNA, but TFIIIA binds not only specifically to the promoter DNA, but also to 5S RNA itself^{6–9}. Increasing evidence indicates that zinc-fingers are more widely used to recognize RNA^{10–13}. There have been numerous structural studies on DNA binding¹⁴, but none on RNA binding by zinc-finger proteins. Here we report the crystal structure of a three-finger complex with 61 bases of RNA, derived¹⁵ from the central regions of the complete nine-finger TFIIIA–5S RNA complex. The structure reveals two modes of zinc-finger binding, both of which differ from that in common use for DNA: first, the zinc-fingers interact with the backbone of a double helix; and second, the zinc-fingers specifically recognize individual bases positioned for access in otherwise intricately folded ‘loop’ regions of the RNA.

The classical zinc-finger motif consists of a sequence of about 30 amino acids containing two histidines, two cysteines and three hydrophobic residues, which are all at conserved positions (Fig. 1b). It forms a small, independently folded domain stabilized by Zn²⁺, which can be used repeatedly in a modular fashion to achieve sequence-specific recognition of DNA³. The domains all have the same structural framework, but achieve chemical distinctiveness through variations in key residues. Structurally, the domain is composed of a β -hairpin and an α -helix pinned together by Zn²⁺. In the canonical Zif268 docking arrangement, the primary contacts are made by the α -helix, which binds in the DNA major groove through primary hydrogen-bond interactions from helical positions –1, 3 and 6 to one strand of the DNA, and through a secondary interaction from helical position 2 to the other strand¹⁴. There are, however, wide variations in this arrangement among the known zinc-finger–DNA complexes.

Thus, DNA binding is well understood, but the molecular basis of recognition of RNA by zinc-fingers has remained elusive. The secondary and tertiary structure of RNA is more complex than that of DNA, comprising internal loops and helical elements closed by hairpin loops. But much has been learned about the binding of TFIIIA to 5S RNA through chemical, biochemical and mutagenesis studies^{16–20}. In particular, the locations of the nine fingers have been mapped with respect to the secondary RNA structure, and also their relative contributions to the overall binding affinity have been determined. The most important region of the 5S RNA is the central half of the molecule comprising loop E, helix V, loop A, helix II and part of helix IV. Accordingly, the most important part of the TFIIIA protein for RNA binding is the central set of fingers 4–7, with the major contribution coming from fingers 4–6 (see Table 1 in ref. 15).

Ideally what is required is an X-ray structure of the complete complex of TFIIIA and 5S RNA, which occurs naturally as a 7S ribonucleoprotein storage particle⁶. But as our attempts to crystallize this, or a reconstituted complex, have failed, we have produced subcomplexes¹⁵ between the essential parts: namely, the central

region of the RNA and fingers 4–6 and/or 4–7. We have checked by RNA footprinting¹⁵ that these three- and four-finger peptides have the same structural relationship to the RNA as they do when they are part of a complete nine-finger protein.

A complex of a truncated 5S RNA with 61 bases (5S RNA 61n) and a three-finger peptide (Fig. 1a, b) diffracted to 3 Å resolution, and we have solved this structure (Fig. 1c). The asymmetric unit contains two complexes of the three-finger peptide and RNA, and an additional peptide molecule with only two of its three fingers ordered. This additional peptide is a 'passenger' in the open crystal structure, making contacts with the outer faces of fingers 4 and 5 in one complex away from the RNA. It is not involved in the specific interaction. The two non-crystallographically related complexes are folded in the same way, with the exception of helix I in the second complex, which is bent slightly differently with respect to the rest of the RNA molecule. Because helix I has no part in the interaction with the three-finger peptide, our conclusions are not affected.

The overall structure of the complex and the geometry of the binding between peptide and RNA is shown in Figs 1d and 2a. As expected from biochemical studies^{16–20}, finger 4 binds to loop E, finger 5 to helix V, and finger 6 to loop A. The hydrogen-bond interactions between peptide and RNA are shown in Fig. 2b. The

distances in Fig. 2c indicate hydrogen bonds, but some of these may be mediated by water, which cannot be determined at the current resolution. The space-filling model (Fig. 1d) shows how the close packed fingers curl around the long axis of the RNA. The spatial relationship (Fig. 2a) between fingers 4 and 5 is similar to that in the complex with DNA²¹, but that between 5 and 6 is different. Both spatial relationships differ from those between fingers in the canonical Zif268–DNA structure and other complexes¹⁴. Allowing for the fact that helix V has 1 base pair less in *Xenopus* 5S RNA than in archaeobacteria, the RNA structure closely resembles that in the 50S ribosomal subunit²² of *Haloarcula marismortui*, as expected from the large amount of conserved sequence.

The structure of loop E is shown in Fig. 3a. It is intricately folded and also structurally rigid, held by various hydrogen bonds and base stacking. The strands, as depicted in the secondary structure, cross over and are stabilized by reciprocal purine stacking: 75G on 99G, and 100A on 77A. Base 75G protrudes outward and is held by a hydrogen bond to 76U, which forms a base triple with A100, as first recognized in an NMR study of loop E (ref. 23). Several of the

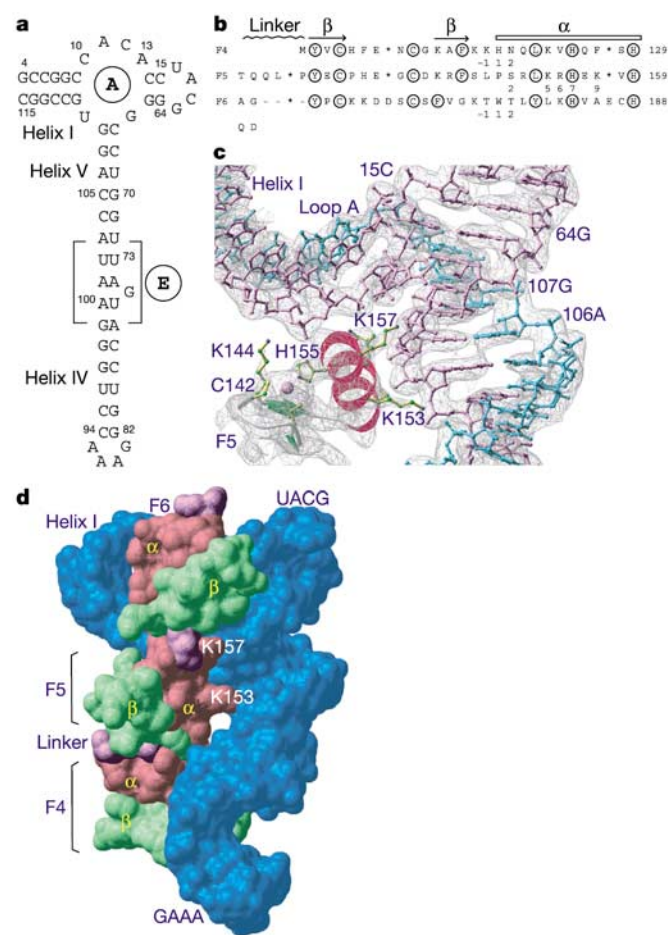


Figure 1 Components of the zinc-finger–RNA complex. **a**, Secondary structure of 5S RNA 61n, the truncated RNA. Loops A and E are circled and helices are indicated by roman numerals. **b**, Sequence of the central three-finger peptide from TFIIIA. Some helical positions are indicated. Conserved residues are circled³, but the alignment in the atypical finger 6 (F6) has been revised according to the structure. **c**, Portion of the electron density map. The RNA is shown as a ball-and-stick model, and finger 5 as a ribbon model. **d**, Space-filling model calculated with a radius of 1.5 Å. RNA is shown in blue, β -sheets in green, α -helices in red, and linkers in purple. The buried surface area is 1,860 Å².

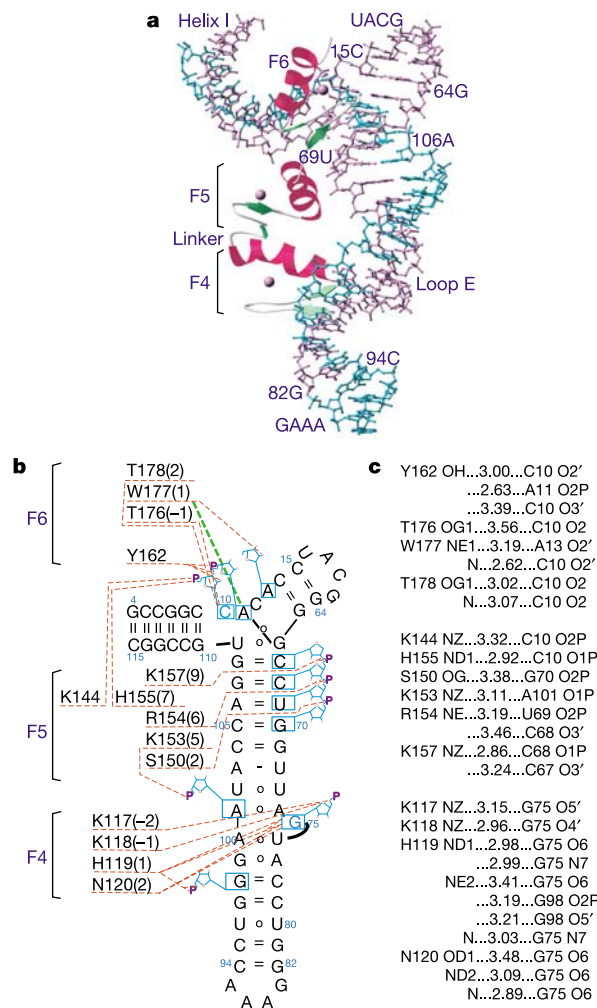


Figure 2 Interactions of the three-finger peptide with the RNA. **a**, Overall view of the complex of TFIIIA(4–6) and 5S RNA 61n. The RNA is represented as a ball-and-stick model with nucleotides 4–82 in purple and nucleotides 83–115 in blue, and the protein is shown as a ribbon model. Purple balls represent zinc ions. **b**, Hydrogen-bond interactions between protein and RNA. Amino acid helical positions in brackets. Nucleotides involved in interactions are boxed, and bases shown in blue make direct interactions with protein. The broken green line indicates stacking of aromatic rings. **c**, List of hydrogen-bond interactions for each of the three fingers.

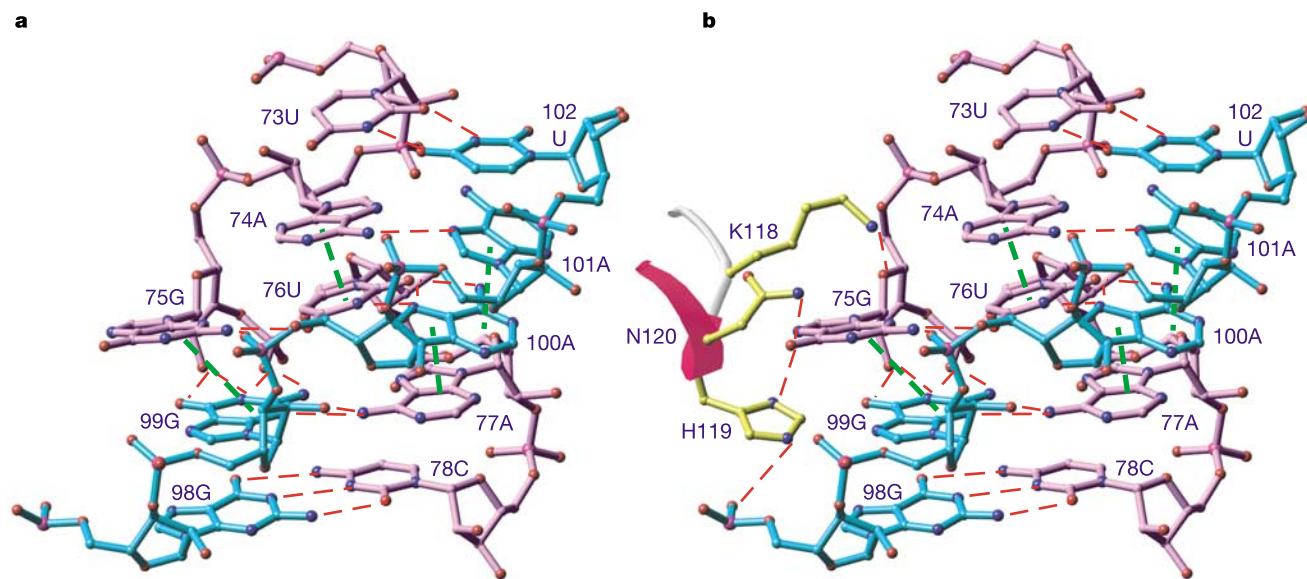


Figure 3 Recognition of loop E by finger 4. **a**, Structure of loop E. The chains are coloured as in Fig. 2a. Hydrogen bonds are shown in red, and base stacking in green. Stacking interactions are assigned according to the degree of overlap and have separation distances shorter than 3.8 Å. **b**, Interaction of loop E with the N terminus of the helix of

finger 4. Colours are the same as in **a**, with peptide side chains in yellow. The hydrogen-bond interactions between protein and RNA are listed in Fig. 2c. The bulged base 75G is gripped by hydrogen bonds from Asp 120 and His 119, and its ribose by a hydrogen bond from Lys 118.

hydrogen-bond interactions were deduced from a detailed mutagenesis study²⁰. This type of structural element has been found in other ribosomal RNAs, where it has been called a “G-bulged cross strand stack”²⁴.

The interactions of loop E with finger 4 are shown in Fig. 3b and listed in Fig. 2c. The bulged base 75G is gripped by hydrogen bonds from Asp 120 and His 119 in positions 2 and 1, respectively, at the amino terminus of the recognition helix, and the ribose of the nucleotide is hydrogen-bonded to Lys 118 in position –1. Thus, the very tip of the helix is used for this specific base recognition. The importance of 75G has long been known because if it is deleted¹⁸, or mutated²⁰, *in vitro* binding to TFIIA is markedly reduced. Quan-

titative mutagenesis studies¹⁹ show that there is a 30-fold reduction in affinity when Lys 118 is replaced by alanine, and a 77-fold reduction when both Lys 118 and Asp 120 are replaced, thereby corroborating our structure.

In terms of the interaction of finger 6 with loop A, what is depicted as a loop in the RNA secondary structure is a complex three-way junction between three RNA helices: V, I and II. At the ends of each of the three helices abutting the loop are two G•C base pairs, which form the rigid pillars of a complex fold (Fig. 4a). The backbone from nucleotides 10–13 follows an extended path, making a sharp ‘S’ bend to join the double helix at nucleotide 14. These four residues form an inner subsidiary loop, closed by the stacking of

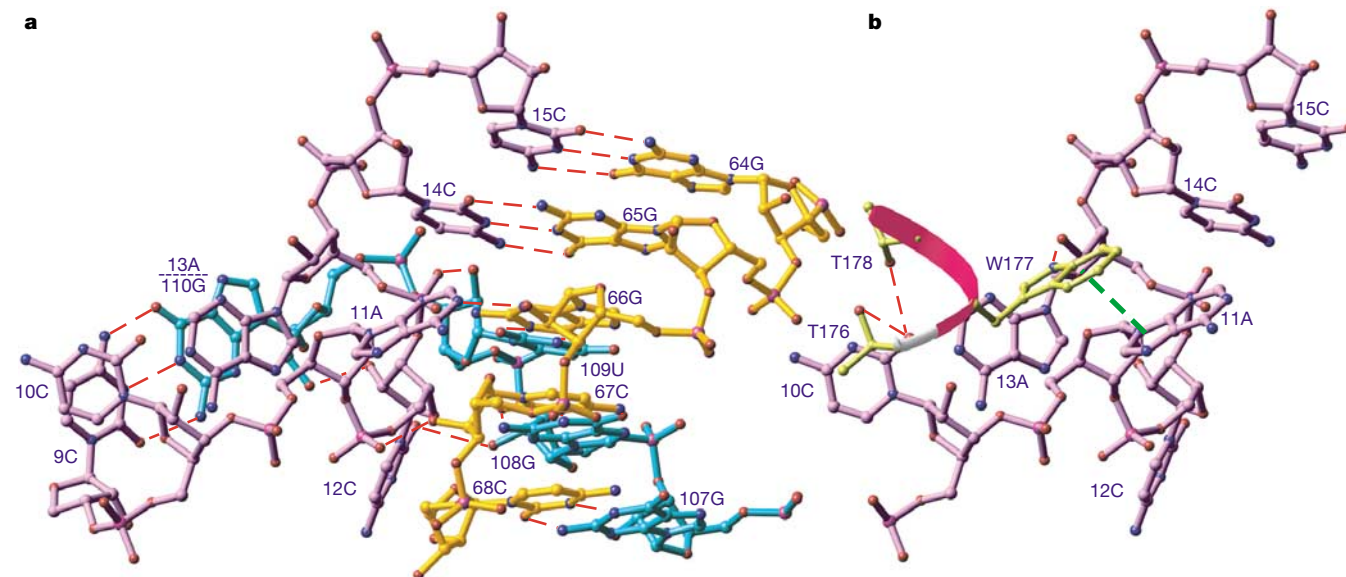


Figure 4 Recognition of loop A by finger 6. **a**, Structure of loop A. Three colours are used to indicate the three-way junction, blue and purple as in Figs 2 and 3, but with nucleotides 64–68 in orange. **b**, Interaction of loop A with the N terminus of the helix of finger 6.

Peptide side chains are shown in yellow. The ring of Trp 177 docks on the face of base 11A, and the two flanking residues, Thr 176 and Thr 178, make hydrogen bonds to base 10C. Trp 177 also makes a hydrogen bond to the ribose of 13A.

13A on 110G, that is hydrogen-bonded to base 9C, which in turn stacks under 10C. Base 11A protrudes outward, fixed by a hydrogen bond to 66G, which forms a base triple with 109U. The face of base 11A is exposed on the outside of the RNA (Fig. 4b).

The binding of finger 6, as in the case of finger 4, involves the N terminus of its recognition helix. Indeed, the same positions at the tip of the helix are involved: namely, -1, 1 and 2, which are occupied here by Thr 176, Trp 177 and Thr 178, respectively (Fig. 4b). The ring of Trp 177 stacks on the face of the exposed base 11A, and the two flanking threonines makes hydrogen bonds to base 10C. Trp 177 also makes a hydrogen bond to the ribose of 13A. Again, there is good agreement with biochemical and mutagenesis studies. The aromatic character of Trp 177, the basis of its stacking, has been reported to be essential for 5S RNA recognition because it can be replaced by phenylalanine but not by alanine²⁰. Of the conserved stretch from 10C to 13A, only one position can be occupied by another base, namely 12C (ref. 20). This is because the base points away from the binding site (Fig. 4b). Alanine substitution of Thr 176 and Thr 178 leads to a 38-fold reduction in binding affinity¹⁹.

The interaction of finger 5 with the RNA differs from the specific base contacts made by fingers 4 and 6. The interaction is made almost completely by multiple contacts of basic amino acid residues with the phosphate backbone (Fig. 2c): two of these, Lys 153 and Lys 157, straddle the major groove of the RNA (Fig. 1c, d). The α -helix of the finger is still used to provide the contacting residues, but these emanate predominantly from the lower part of the helix towards the C terminus. Support for this conclusion comes from alanine mutagenesis of the helical positions involved in DNA binding, which show little effect¹⁹. Strong corroborating evidence for the involvement of the backbone phosphates comes from ethylation interference assays²⁰. Indeed, it is remarkable that these previous biochemical studies succeeded in identifying many of the essential moieties in both protein and RNA.

The interaction of finger 5 with the double-stranded RNA of helix V is distinct from that with the DNA double helix²¹. The base sequence of RNA helix V is identical (except for U replacing T) with that of the parent DNA double helix, so that finger 5 might have been expected to bind in a similar manner. This does not happen, presumably because the major groove of the RNA double helix is too deep to be penetrated by the α -helix. The α -helix is still used and provides nonspecific contacts to the phosphate sugar backbone. Presumably, the register of the run of fingers with the RNA is decided by the specific base contacts made by fingers 4 and 6, restricting finger 5 to bind to the RNA double helix in between. Finger 5 therefore seems to be a nonspecific spacer element that, through backbone contacts, contributes substantially to the binding affinity. This situation is reciprocal to the way in which the three-finger peptide associates with DNA²¹, where finger 5 makes specific contacts and fingers 4 and 6 act as non-binding spacers.

The dual role of finger 5 binding to DNA and RNA in different ways points to the wide difference between these molecules. The primary interaction of fingers with DNA is base recognition. The DNA double helix is essentially regular, despite small, local, sequence-dependent variations, thus the specificity of recognition lies in a particular sequence of bases. By contrast, RNA molecules form complex structures comprising internal 'loops' and double helices, which are sometimes closed by hairpin loops. The general conclusion from our work is that the zinc-finger domain can be used to recognize all of these different elements. It can recognize RNA specifically by contacts with individual bases that are exposed for access out of a structurally rigid, complicated fold. At the same time, the zinc-finger architecture is versatile enough to provide binding in another mode to a regularly folded double-helical region, where it is the structure rather than the sequence that is recognized. Thus, the finger 5 interaction may account for other instances of specific binding by zinc-fingers to double-stranded RNA that is not

sequence dependent^{11,12}. It will be interesting to compare zinc-finger usage more generally with other types of RNA-binding protein. A diverse set of structures is known, but so far there is no comprehensive library of interactions, including those from the structures of the ribosomal units.

The adaptability of the zinc-finger for molecular recognition is further attested to by increasing evidence for the use of zinc-fingers in protein-protein recognition; for example, the six-finger protein Aiolos uses two of its fingers to bind to another zinc-finger Ikaros and the remaining four for binding DNA²⁵. In addition, it seems that zinc-fingers are still rapidly evolving², presumably because, as small self contained modules, they can be easily passed on in exon shuffling. There is more to be learned. □

Methods

Reconstitution of zinc-finger-RNA complex

RNA and protein were prepared as described¹⁵. RNA and protein samples were always freshly refolded and trial binding was carried out before the bulk reconstitution of complex. Binding affinity can vary slightly between different batches of protein: the molar ratio of RNA to protein is 1:1 normally and 1:2 in extreme cases. Extra protein was added in bulk preparations to secure 100% of bound RNA.

We mixed the RNA and protein in dilute concentrations ($\sim 4.7 \mu\text{M}$) to avoid irreversible precipitation. The mixture was then dialysed against binding buffer (20 mM HEPES pH 7.0, 150 mM KCl, 1.5 mM MgCl_2 , 5 mM dithiothreitol (DTT) and 0.1 mM ZnSO_4) at 4°C overnight. A white precipitate of aggregated protein normally appeared and was removed by ultracentrifugation at 55,000 r.p.m., 4°C, 30 min. The supernatant was concentrated in a Vivaspin concentration unit (Vivascience) to a final concentration of 5 mg ml⁻¹.

Crystallization

The 5S RNA 61n construct was designed with two different tetraloops capping the stem loops. The particular choice of the two tetraloops (Fig. 1a) differs from that shown in Fig. 4 of ref. 15, because it was found to be essential for crystallization of the complex.

Crystallization was carried out by the hanging-drop evaporation method at 4°C. Concentrated complex was mixed with reservoir buffer (20% PEG 8,000, 200 mM KCl, 5 mM MgCl_2 , 50 mM MES, (pH 5.6), 3 mM DTT and 0.3 mM ZnSO_4) at a 1:1 volume ratio. Flexible needles normally appeared in 2 d. Very occasionally, block-shaped crystals appeared after 2 weeks. These were used for seeding, and block-shaped crystals were reproduced after 2–3 d. We checked the composition of crystal batches by gel electrophoresis.

Data collection

Crystals were soaked in increasing steps of 30 min each into cryo-protection buffer containing 18% ethylene glycol: 5% in the first three steps and 3% in the last step. A crystal held in a loop was frozen by dipping it directly into liquid nitrogen.

The crystals diffracted close to 3 Å resolution in space group C2, with cell dimensions of $a = 58.6 \text{ Å}$, $b = 191.6 \text{ Å}$, $c = 79.8 \text{ Å}$ and $\beta = 101.5^\circ$. Two-wavelength multiple anomalous dispersion (MAD) data sets at the zinc absorption edge were collected at beam line BM30A of the European Synchrotron Radiation Facility (ESRF), Grenoble. The data set at the peak wavelength of 1.28200 Å was collected to 3.2 Å, and that at the inflection wavelength of 1.28347 Å was collected to 3.3 Å. Subsequently, we collected a slightly higher resolution data set to 3.1 Å at beam line ID14.2 of the Synchrotron Radiation Source, Daresbury, UK.

Structure determination

Data sets were processed with MOSFLM and scaled with SCALA in the CCP4 suite²⁶. Data sets collected at the peak wavelength and the inflection wavelength, and the high resolution gave an R_{merge} of 6.9%, 6.6% and 5.1%, respectively. The first zinc sites were obtained by SHELX²⁷ and were refined in SHARP²⁸. SHARP produced a clear experimental phasing map with a solvent content of 50%. This showed in the asymmetric unit the presence of two complexes and one extra protein molecule with only two of its fingers ordered. We built a model of one of the complexes manually using the program O (ref. 29), and searched out the second complex out by using the model of the first complex with the program MOLREP in the CCP4 package²⁶. The extra two ordered fingers were built manually in O. The whole asymmetric unit was first refined with Refmac5 in the CCP4 package²⁶ to reach values of $R_{\text{free}} = 30\%$ and $R_{\text{work}} = 20\%$; the refinement was then transferred to CNS³⁰ to reach final values of $R_{\text{free}} = 25.6\%$ and $R_{\text{work}} = 21.5\%$ (r.m.s. bond length, 0.0067 Å; and r.m.s. bond angle, 1.27°). The Ramachandran plot for the protein shows six residues in the generously allowed regions and only one in the disallowed regions. Figures of the models were drawn in Ribbons.

To check that the extra peptide had no effect on the complex, we solved the structure of another complex with 57 bases of RNA, in which the GAAA tetraloop was removed at the end of helix IV and which produced crystals that diffracted to 3.5 Å resolution. This crystal grew in a different but related space group of C222, with a similar cell volume. The asymmetric unit contained only one complex, and the structure was solved by molecular replacement to a value of $R_{\text{free}} = 0.4$, but not refined. This showed the same arrangement as the current structure without the extra peptide in first crystal, but we did not proceed with the refinement because of the lower resolution.

Received 1 August; accepted 22 September 2003; doi:10.1038/nature02088.

1. Clamp, M. *et al.* Ensembl 2002: accommodating comparative genomics. *Nucleic Acids Res.* **31**, 38–42 (2003).
2. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **30**, 276–280 (2002).
3. Miller, J., McLachlan, A. D. & Klug, A. Repetitive zinc-binding domains in the protein transcriptional factor IIIA from *Xenopus* oocytes. *EMBO J.* **4**, 1609–1615 (1985).
4. Sakonju, S. & Brown, D. D. Contact points between a positive transcription factor and the *Xenopus* 5S RNA gene. *Cell* **31**, 395–405 (1982).
5. Engelke, D., Ng, S.-Y., Shastry, B. & Roeder, R. Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. *Cell* **19**, 717–728 (1980).
6. Picard, B. & Wegnez, M. Isolation of a 7S particle from *Xenopus laevis* oocytes: a 5S RNA–protein complex. *Proc. Natl Acad. Sci. USA* **76**, 241–245 (1979).
7. Pelham, H. & Brown, D. A specific transcription factor that can bind to either the 5S RNA gene or 5S RNA. *Proc. Natl Acad. Sci. USA* **77**, 4170–4174 (1980).
8. Bogenhagen, D. F. & Sands, M. S. Binding of TFIIIA to derivatives of 5S RNA containing sequence substitutions or deletions defines a minimal TFIIIA binding site. *Nucleic Acids Res.* **20**, 2639–2645 (1992).
9. Theunissen, O., Rudt, F., Guddat, U., Mentzel, H. & Pieler, T. RNA and DNA binding zinc fingers in *Xenopus* TFIIIA. *Cell* **71**, 679–690 (1992).
10. Nagai, K. *et al.* Structure and assembly of the spliceosomal snRNPs. *Biochem. Soc. Trans.* **29**, 15–26 (2001).
11. Finerty, P. J. & Bass, B. L. A *Xenopus* zinc finger protein that specifically binds dsRNA and RNA–DNA hybrids. *J. Mol. Biol.* **271**, 195–208 (1997).
12. Mendez-Vidal, C., Wilhelm, M. T., Hellborg, F., Qian, W. & Wiman, K. G. The p53-induced mouse zinc finger protein wig-1 binds double-stranded RNA with high affinity. *Nucleic Acids Res.* **30**, 1991–1996 (2002).
13. Ladomery, M., Somerville, J., Woolner, S., Slight, J. & Hastie, N. Expression in *Xenopus* oocytes shows that WT1 binds transcriptions *in vivo*, with a central role for zinc finger one. *J. Cell Sci.* **116**, 1539–1549 (2003).
14. Wolfe, S. A., Neklodova, L. & Pabo, C. O. DNA recognition by Cys₂His₂ zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 183–212 (2000).
15. Searles, M. A., Lu, D. & Klug, A. The role of the central zinc fingers of transcription factor IIIA in binding to 5S RNA. *J. Mol. Biol.* **301**, 47–60 (2000).
16. Clemens, K. R. *et al.* Molecular basis for specific recognition of both RNA and DNA by a zinc finger protein. *Science* **260**, 530–533 (1993).
17. McBryant, S. J. *et al.* Interaction of the RNA binding fingers of *Xenopus* transcription factor IIIA with specific regions of 5S ribosomal RNA. *J. Mol. Biol.* **248**, 44–57 (1995).
18. Setzer, D. R., Menezes, S. R., del Rio, S., Hung, V. S. & Subramanyan, G. Functional interactions between the zinc fingers of *Xenopus* transcription factor IIIA during 5S rRNA binding. *RNA* **2**, 1254–1269 (1996).
19. Friesen, W. J. & Darby, M. K. Phage display of RNA binding zinc fingers from transcription factor IIIA. *J. Biol. Chem.* **272**, 10994–10997 (1997).
20. Theunissen, O., Rudt, F. & Pieler, T. Structural determinants in 5S RNA and TFIIIA for 7S RNP formation. *Eur. J. Biochem.* **258**, 758–767 (1998).
21. Nolte, R. T., Conlin, R. M., Harrison, S. C. & Brown, R. S. Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex. *Proc. Natl Acad. Sci. USA* **95**, 2938–2943 (1998).
22. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920 (2000).
23. Wimberley, B., Varani, G. & Tinoco, I. The conformation of loop E in eukaryotic 5S ribosomal RNA. *Biochemistry* **32**, 1078–1087 (1993).
24. Correll, C. C. *et al.* Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proc. Natl Acad. Sci. USA* **95**, 13436–13441 (1998).
25. Morgan, B. *et al.* Aiolos, a lymphoid restricted transcription factor that interacts with Ikaros to regulate lymphocyte differentiation. *EMBO J.* **16**, 2004–2013 (1997).
26. Collaborative Computational Project 4, The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
27. Sheldrick, G. M. & Gould, R. O. Structure solution by iterative peaklist optimization and tangent expansion in space group P1. *Acta Crystallogr. B* **51**, 423–431 (1995).
28. de la Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
29. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
30. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

Acknowledgements We acknowledge the Daresbury Laboratory and the ESRF for provision of facilities. We thank P. Evans and other colleagues for advice and help. D.L. was initially supported by a grant from the Human Frontier Science Programme (to A.K.) and later by a Fellowship from the Sino-British Fellowship Trust.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.K. (akl@mrc-lmb.cam.ac.uk). Coordinates are deposited in the Protein Data Bank under accession code 1UN6.

retraction

Hes1 is a target of microRNA-23 during retinoic-acid-induced neuronal differentiation of NT2 cells

Hiroaki Kawasaki & Kazunari Taira

Nature **423**, 838–842 (2003).

In this Article, the messenger RNA that is identified to be a target of microRNA-23 (miR-23) is from the gene termed human ‘homolog of ES1’ (HES1), accession number Y07572, and not from the gene encoding the transcriptional repressor ‘Hairy enhancer of split’ HES1 (accession number NM_00524) as stated in our paper. We incorrectly identified the gene because of the confusing nomenclature. The function of HES1 Y07572 is unknown but the encoded protein shares homology with a protein involved in isoprenoid biosynthesis. Our experiments in NT2 cells had revealed that the protein levels of the repressor Hes1 were diminished by miR-23. Although we have unpublished data that suggest the possibility that miR-23 might also interact with Hes1 repressor mRNA, the explanation for the finding that the level of repressor Hes1 protein decreases in response to miR-23 remains undefined with respect to mechanism and specificity. Given the interpretational difficulties resulting from our error, we respectfully retract the present paper. Further studies aimed at clarifying the physiological role of miR-23 will be submitted to a peer-reviewed journal subject to the outcome of our ongoing research. □

addendum

An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene

Shinji Hirotsune, Noriyuki Yoshida, Amy Chen, Lisa Garrett, Fumihito Sugiyama, Satoru Takahashi, Ken-ichi Yagami, Anthony Wynshaw-Boris & Atsushi Yoshiki

Nature **423**, 91–96 (2003).

In this Letter, it is shown using transgene insertion mouse mutants that the *Makorin1-p1* pseudogene regulates the expression of its related coding gene. An example has been drawn to our attention of another transcribed pseudogene that regulates the expression of its related coding gene, but by a different mechanism, in the mollusc *Lymnaea stagnalis*¹. □

1. Korneev, S. A., Park, J.-H. & O’Shea, M. Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. *J. Neurosci.* **19**, 7711–7720 (1999).

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Frank Phelan (4944)
**Netherlands/ Italy/ Spain/
Portugal/ Belgium:**
Evelina Rubio Hakansson (4973)
Scandinavia: Silje Opstrup (4994)
France/ Switzerland:
Amelie Pequignot (4974)
Natureevents:
Paul Constant (4954)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria:
Patrick Phelan, Odo Wulffen
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com
o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
Asia-Pacific Sales Director:
Rinoko Asami
e-mail: rasami@naturejpn.com

naturejobs

Trust me, I'm a doctor

How many advanced degrees does one really need? PhD/JDs are growing. PhD/MBAs abound. And MD/PhDs are downright commonplace. A panel of young scientists-turned-businesspeople debated the question last week at the Yale Biotechnology Student Interest Group in New Haven, Connecticut.

Julie Huang started out studying for an MD, moved to public health and, after completing her MPH, "followed the herd to Wall Street" to work as an investment banker before switching to biomedical public relations. She is now vice-president of public-relations agency Cohn & Wolfe Healthcare in New York.

After a postdoc at Johns Hopkins University, John Puziss joined chemical firm Bristol-Myers Squibb, where he started out in fungicides but ended up looking for technology partners. He turned that experience, and an encounter with a postdoc-era friend, into a business-development job at a small proteomics firm. But as the company grew, he found himself doing more sales. So he jumped at the chance to work in technology transfer at Yale University, where he is associate director of its Office of Cooperative Research.

Frank Wang realized during his fellowship that he didn't want to take the academic track. Landing a job with London-based international consultancy McKinsey, he spent a year each in Beijing and Hong Kong. Then the economy faltered and the company shed employees; he is now on Yale's MBA course.

Wang thinks an MBA will make him more competitive. Puziss is not convinced that more degrees are the answer for everyone. Huang concludes that there is no formula, and no easy way to reach career goals, as each story illustrates: "You have to create your own path, as crazy as it might seem or as difficult as it might be."

Paul Smaglik

Naturejobs editor



Contents

POSTDOCS AND STUDENTS

A helping hand for grant proposals p102

CAREER VIEW

Nuts & Bolts

Networking for success

Young Scientist

Fighting for postdoc subsidies

Movers

Per Ahlberg p104

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

A winning proposal

Young, aspiring researchers often have to learn the hard way when it comes to writing a killer grant application. But a range of European initiatives aims to give them a helping hand. Karen Kreeger reports.

Jonas Ludvigsson wrote nine applications before finally getting a grant to finance his first laboratory study. “That was a tough time,” recalls the paediatrics postdoc at Örebro University Hospital in Sweden.

Most students and postdocs, like Ludvigsson, learn grantsmanship through trial and error. But learning the ropes alone is hardly a recipe for instant success. And until relatively recently, there weren’t many solutions available to European students beyond hoping that their principal investigator would offer some useful advice — making the choice of mentor a critical one (see *Nature* **422**, 784–785; 2003).

This situation is beginning to change. University-based workshops are emerging, as is better training through foundations such as the Wellcome Trust, and clearer instruction from funding bodies like the European Commission.

LEARN FROM EXPERIENCE

Ludvigsson’s eight attempts before his ultimate success were not entirely in vain. He turned the material he gathered from a variety of sources (see Web links, opposite) into a career-development course, which he taught for the first time this year.

Katja Hagström, a doctoral student in occupational medicine also at Örebro University Hospital, says that the course taught her to keep her proposal straightforward and easy to read. And, perhaps most importantly, Hagström learned to emphasize in the application why her research is significant.

Walter Osika, a doctoral student in cardiovascular research at Örebro University Hospital who also took Ludvigsson’s course, says that novice grant-writers should read the instructions closely and not leave their proposal until the last minute. He also recommends that students show early drafts not just to their supervisor but also to friends or family — after all, if non-researchers cannot understand the basic ideas, the proposal is probably unclear and overly technical.

Hans-Olov Adami, from the department of medical epidemiology and biostatistics



Hans-Olov Adami feels that grant-writing is a neglected area in European institutions.

at the Karolinska Institute in Stockholm, Sweden, says that classes in grant-writing are a neglected area in European institutions. “To compensate for that, we have a long tradition in our department to organize courses in grant-writing for our senior graduate students,” he says. These semester-long courses routinely end with participants submitting a research grant proposal, usually to a Swedish funding source.

TIPS FROM THE TOP

Some charities, funding agencies and private consultancies also impart grant-writing advice to aspiring researchers. Mary Phillips, scientific programme manager at Britain’s Wellcome Trust, the world’s largest biomedical charity, says that the trust largely gives grant-writing advice on a case-by-case basis, usually to researchers who are unfamiliar with its application procedure. She has written a short article that is sent out with application forms and distributed at meetings. It covers ‘signposts’ for writing a competitive application (see ‘Wellcome signposts’, left).

The European Commission’s Sixth Framework Programme devotes E1.6 billion (US\$1.9 billion) — almost one-tenth of its budget — to promoting training, mobility and career development for researchers through its Marie Curie Actions (see *Nature* **423**, 98–99; 2003). Estelle Kane, the commission’s UK contact for human resources and mobility, regularly holds information days on applying for the fellowships. Those sessions emphasize understanding the evaluation process and the importance of meeting the scheme’s deadlines.

Malene Cortelius, a consultant for the Sixth Framework Programme based at the EuroCenter, part of the Danish Technological Institute in Taastrup, has teamed up with Hyperion, a consulting firm run

Wellcome signposts

Mary Phillips, scientific programme manager at UK biomedical charity the Wellcome Trust, has written a short article for the benefit of those who are applying for the first time for one of the trust’s grants. But the basic advice that she sets out in the article, “Grantmanship: signposts for competitive grant applications”, could act as a checklist for any grant hopeful. Here are a few of the questions that grant-writers should ask themselves before they submit an application:

- Is the work novel, exciting and necessary?
- Are you repeating experiments that have already been undertaken?
- Have you justified various aspects of the funding requested?
- Have you filled out the ‘housekeeping’ portions of the form correctly?
- Have you checked your form for spelling and grammatical mistakes? **K.K.**

♦ www.wellcome.ac.uk



Teachers: Shaun McCarthy (above) and Malene Cortelius have teamed up to coach European researchers in maximizing their chances of securing a grant.



by Shaun McCarthy, an Irish engineer who has been the recipient of more than 150 European grant proposals, to teach grant-writing skills for European project-management courses and Framework grants. The course provides guidelines for dealing with the complexity of the multi-country, multi-discipline proposals, focusing particularly on how to provide details of collaborators and budgets.

Lizzie Melby Jespersen, international projects manager at the Danish Institute of Agricultural Sciences (DIAS) in Tjele, has attended several of Hyperion's courses and has enlisted McCarthy to teach in-house courses at the institute on grant-writing and contract negotiation for postdoctoral fellows, advanced researchers, research administrators and heads of research departments. Jespersen says that such formal training is effective. Since working with Hyperion, DIAS's success rate in obtaining grants from the European Commission has risen considerably.

Sieglinde Gruber, who works for the European Commission's Strategy and Policy Aspects Unit in Brussels, notes that the Marie Curie Actions do not specifically provide grant-writing advice to postdocs and advanced students. But she agrees that such skills are part of the broader delivery of training through research, which is the main objective of some of the Marie Curie Actions.

MAXIMUM MENTORING

Formal grant-writing training is "an interesting topic", says Luc Van Dyck, executive coordinator of the European Life Sciences Forum in Heidelberg, Germany. "But I think it still remains mostly a business between supervisors and trainees," he argues.

The problem is that many senior researchers don't involve their junior colleagues in grant-writing early

enough in their careers, says Olof Akre, a researcher in the Karolinska Institute's clinical epidemiology unit. If informal approaches fail, more formal methods become necessary, he believes.

That's why Akre, along with his colleague Adami and Anders Ekbom, a professor in the department of medicine at Karolinska Hospital, have come together to teach a grant-writing course — as well as encouraging their own students to apply for grants early and often. The course requires students to complete at least one grant application.

Ekbom stresses the importance of good writing in producing a winning grant proposal. "A well-written application is more likely to be funded than a badly written one," he says. "It may sound simplistic, but it is a hard message to convey."

To make this message stick, the trio ends the course by having the students form review boards to judge the quality of each others' applications, and rank them in order of strength, just as a real granting body would. "This is a painful process, but grantsmanship has to be learned — no one is born a grant-writer," Ekbom says.

Gunnar Larfors, one of Ekbom's graduate students, has written successful grants with his mentor. But the formal course demystified the process and made him more confident about going it alone. "Being able to raise your own money is a crucial step in the path to independence as a scientist," Larfors says.

Whatever the method of education — learning from a mentor, taking a formal course or educating yourself through materials supplied by funding agencies — the key is to take the first step early, and not wait until the end of your doctoral or postdoctoral work to learn the ropes of securing funding. A little education and preparation will take some of the error out of this 'trial-and-error' process.

Karen Kreeger is a freelance writer based in Media, Pennsylvania.

Web links

How to write a grant application (US National Institute of Allergy and Infectious Diseases)

♦ www.niaid.nih.gov/ncn/grants/write/write.pdf
Grantslam

♦ www.cayuse.com/productinfo.htm
Marie Curie Actions

♦ europa.eu.int/mariecurie-actions
EU Mobility Portal

♦ www.ukro.ac.uk/public/Mobility/Events/ncpeventautumn.htm
Sixth Framework

♦ fp6.cordis.lu/fp6/home.cfm

Hyperion Training Courses for Research Managers

♦ www.hyperion.ie

YOUNG SCIENTIST

Paying the rent

A plan to phase out housing subsidies for postdocs at Rockefeller University in New York has been put on hold. Suggested in July, the move would have slowly reduced a 15% rent subsidy to zero.

The proposal was put forward long before new university president Paul Nurse was due to take up his post full-time; and when the university's postdoctoral association learned of the plan, it mounted a concerted effort to tell him what the financial impact of the cut would be. The subsidy, it said, was what made it possible for postdocs surviving on a stipend to afford to live in New York.

Reacting to the postdocs' concerns, Nurse has suspended any changes pending a comprehensive review of the entire postdoc remuneration package, including salary, housing, health benefits and day care — all issues raised in a recent survey that was commissioned by the postdoctoral association.

The challenge now is for the whole Rockefeller community to work together to resolve this situation. But if the tone of the welcoming party held by the postdoctoral association for the new president is any indication, Rockefeller postdocs can confidently look forward to a fruitful existence under his guidance. ■

The Rockefeller University Postdoctoral Association

♦ www.rockefeller.edu/pda

NUTS & BOLTS

Dale Carnegie, who wrote the best-selling book *How to Win Friends and Influence People*, asserts that you can make more contacts in two months by becoming interested in others' work than by spending two years trying to get them interested in yours.

But how can this help your science career? Simple. Developing a network of mutually beneficial contacts can ease any of your career challenges — from changing job, to securing a grant or getting a promotion recommendation.

Networking is a give-and-take process that requires a genuine interest in and focus on others. It is not limited to the staged and formal gatherings of experts; it happens naturally all the time. Everyone, even a social acquaintance outside your profession, is a potential resource who may know someone or something that will prove useful to your



With Deb Koen
Careers consultant

career. Move beyond your usual sphere of influence to diversify your network of people and ideas. Start with a fairly comfortable venue and progress to more personally challenging situations.

To ease into first-time conversations, develop a three-part introduction that includes your name, your speciality or focus, and an interesting point of distinction. If you can get past the introduction and are a good listener, you'll be ready on any occasion with a flowing response to the recurring question: "So, what do you do?" If you are

seeking assistance beyond the introduction, be concise and specific in your request. Get to the point, be considerate of time, and listen as much as you speak.

Remember, you need to add value as well as make requests. So consider the support, information and resources that you can contribute. Help others by reviewing manuscripts for journals, serving on study sections for granting bodies, taking part in professional associations, speaking on alumni panels, adopting a collaborative approach at work and sharing credit with those who have helped you.

Networking is a vital force in both getting work and getting work done. To enjoy the benefits, steer clear of the 'me, me, me' mindset and focus on what you can do for others. ■

Deb Koen is vice-president of Career Development Services, a national career-management firm, and is a columnist for *The Wall Street Journal's CareerJournal.com*.

MOVERS

Per Ahlberg, Professor of Evolutionary Organismal Biology, Uppsala University, Sweden



Per Ahlberg's career as a palaeontologist has been marked by a constant search for balance — between teaching and research, his professional and personal life, and his native Sweden and his adopted home in England. His new position at Uppsala University, he says, brings all of those elements into line.

After earning his PhD in 1989, Ahlberg took up a post, the title of which is now extinct — he became a 'departmental demonstrator' at the University of Oxford. That meant a fixed-term contract, with a heavy load of supervising practical lab classes and individual tutorials. Looking

back, Ahlberg feels that the mix of teaching and research stood him in good stead. But at the time, he was worried that the position would shackle him professionally, as his teaching commitments meant he couldn't take on the kind of research that can make or break a palaeontologist's career.

Ahlberg used a combination of fortune, experience, knowledge and instinct to turn this disadvantage on its head. Because he had little time or budget for travel, he devoted much of his non-teaching time to combing through collections in museums. At one museum, he noticed that some fossils that had been classified as bony-fish fragments in the nineteenth century actually had features, such as a distinctive snout, associated with land-based vertebrates, or tetrapods, that he had collected on a postgraduate dig in Greenland. Ahlberg showed that the mislabelled bony fish was actually

the oldest known fossil of a tetrapod. The discovery served as a springboard for his scientific career and reinforced his inclination never to accept conventional wisdom.

A series of related findings led to a post at London's Natural History Museum. But after nearly a decade, the pendulum had swung too far in the other direction. He had plenty of time for research and travel, but was missing teaching. "I like the contact with the undergrads," Ahlberg says. And the 90-minute journey to work in central London took a toll, particularly after the birth of his daughter in 2001.

The Uppsala appointment resolves all those issues, he says. He has the right teaching load now, and the office is a five-minute bike ride from home. Ahlberg sums up his career with a simple statement: "I was able to recognize some things that hadn't been recognized." The same might be said for his recognition of the need for balance. ■

CV

1994–2003: Researcher (fossil fish and amphibians), palaeontology department, Natural History Museum, London

1989–1994: Departmental demonstrator (vertebrate biology), zoology department, University of Oxford

1989: PhD in Zoology, University of Cambridge